

OM of: US-09-303-518D-125 to: SPREMBL_19.* out-format: pfs
 Date: Jun 30, 2002 8:00 AM

About: Results were produced by the Gencore software, version 4.5.
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framed.n2p.model -DEV=xlh
 -O=/cgn2.1/US9303518/unat_28062002.14271.4338/app-query.fasta.1.23501
 -DB=SPREMBL_19 -OPMT=fastan -SUFFIX=tspt -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPECL=0.000 -LOOPEXT=0.000
 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
 -DELDP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsu62
 -TRAP=human40.cdi -LIST=100 -DOCALLIGN=200 -THR_SCORE=Pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTPRF=Pfs
 -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
 -USER=US9303518 -CGN1_1.1967 -NCPY=6 -ICPY=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-303-518D-125
 Query length: 1344
 Database: SPREMBL_19.*
 Database sequences: 562222
 Database length: 172994929
 Search time (sec): 883.700000

Sequence	Strd Orig	Zscore	Escore	Len	Documentation
SP_bacteria:044130	1567.00	264.01	7.7e-113	449	I 044130 actinobacillus pleuropne
SP_bacteria:09X650	663.00	908.05	6.9e-43	451	I 09X650 porphyromonas gingivalis
SP_bacteria:09W86	140.00	181.45	0.0205	451	I 09W86 thermotoga maritima, ele
SP_fungi:004051	135.00	159.77	0.0828	1802	I 004051 saccharomyces cerevisia
SP_fungi:09C105	132.00	159.61	0.1232	1262	I 09C105 schizosaccharomyces pom
SP_human:060382	131.00	159.61	0.1232	1262	I 060382 homo sapiens (human). X
SP_human:09U035	131.00	149.71	0.1971	2752	I 09U035 homo sapiens (human). X
SP_human:09U036	130.00	156.50	0.1780	1275	I 09U036 homo sapiens (human). X
SP_human:015038	130.00	152.94	0.2011	1783	I 015038 homo sapiens (human). X
SP_fungi:09A317	126.00	160.20	0.2644	534	I 09A317 schizosaccharomyces pomt
SP_bacteria:09B8E3	120.50	150.54	0.7551	646	I 09B8E3 rhizobium loti (mesorhiz
SP_human:09H0M1	119.50	149.26	0.8988	639	I 09H0M1 homo sapiens (human). X
SP_invertebrate:09VXMS	119.50	135.16	1.46	2406	I 09VXMS drosophila melanogast
SP_plant:09S0V0	118.00	144.06	1.31	857	I 09S0V0 arabidopsis thaliana (mc
SP_bacteria:0931B1	115.50	151.78	1.13	299	I 0931B1 bifidobacterium animalis
SP_human:09U83	114.50	143.35	2.11	580	I 09U83 homo sapiens (human). hy
SP_rodent:09Y35	114.50	130.29	3.31	1979	I 09Y35 mus musculus (mouse). h
SP_mammal:028071	114.50	142.96	2.44	528	I 028071 sus scrofa (pig). gastri
SP_bacteria:093S58	112.00	139.97	3.29	575	I 093S58 myxococcus xanthus. usc3
SP_rodent:09XK60	112.00	138.98	3.40	631	I 09XK60 mus musculus (mouse). mu
SP_invertebrate:09U0V2	112.00	135.66	5.37	2207	I 09U0V2 leishmania major. pos
SP_bacteria:09RKP2	111.50	143.79	3.08	376	I 09RKP2 deionococcus radiodurans
SP_human:09U039	111.50	130.91	4.79	956	I 09U039 homo sapiens (human). r
SP_bacteria:09U040	111.50	139.46	3.82	1262	I 09U040 homo sapiens (human). X
SP_bacteria:09A909	111.00	136.32	4.25	529	I 09A909 microbacterium ammoniaph
SP_vertebrate:091BCT	111.00	136.32	4.25	711	I 091BCT xenopus laevis (afriean
SP_invertebrate:09N3Y8	110.50	122.94	7.17	2344	I 09N3Y8 caenorhabditis elegans
SP_invertebrate:091A95	110.00	124.78	7.19	1847	I 091A95 caenorhabditis elegans
SP_bacteria:092A63	109.50	144.53	3.90	270	I 092A63 neisseria meningitidis
SP_bacteria:051099	109.50	133.35	5.72	773	I 051099 borrelia burgdorferi (l)
SP_invertebrate:022579	109.50	124.10	7.85	1844	I 022579 caenorhabditis elegans
SP_invertebrate:091365	109.50	122.07	8.42	2232	I 091365 caenorhabditis elegans
SP_bacteria:09KXT0	109.00	130.35	6.77	960	I 09KXT0 streptomyces coelicolor
SP_human:09U8A8	108.50	120.38	10.17	2296	I 09U8A8 homo sapiens (human). s
SP_bacteria:093RM3	108.00	138.11	5.91	406	I 093RM3 streptomyces coelicolor
SP_rodent:061191	108.00	120.91	10.65	2045	I 061191 mus musculus (mouse). d
SP_rodent:09Q0M2	108.00	120.91	10.65	2045	I 09Q0M2 mus musculus (mouse). d
SP_virus:096BM2	107.50	123.49	10.41	1504	I 096BM2 avian leukosis virus. g
SP_bacteria:097PQ2	107.00	128.09	9.50	914	I 097PQ2 streptococcus pneumoniae
SP_human:095769	107.00	128.08	9.50	915	I 095769 homo sapiens (human). wu

SP_invertebrate:076602	-	107.00	124.55	10.72	1275	I 076602 caenorhabditis ele
SP_bacteria:005089	+	106.50	132.90	8.60	545	I 005089 nocardioideus simplex
SP_human:096H41	+	106.50	126.61	10.66	984	I 096H41 homo sapiens (human).
SP_archaeap:097A73	+	106.00	131.32	9.69	592	I 097A73 thermoplasma volcani
SP_invertebrate:09BK7	-	106.00	122.75	12.99	1325	I 09BK7 leishmania major.
SP_invertebrate:09U021	-	106.00	114.50	17.24	2879	I 09U021 leishmania major.
SP_virus:091TW8	+	105.50	129.15	11.14	419	I 091TW8 maize rayado flavo vir
SP_bacteria:052747	+	105.50	124.30	9.34	680	I 052747 runioccoccus flavefac
SP_invertebrate:09V8K9	+	105.50	118.18	16.22	1908	I 09V8K9 drosophila melanog
SP_human:09H8A8	+	105.50	116.21	17.35	2296	I 09H8A8 homo sapiens (human)
SP_rodent:09QX47	+	105.50	115.72	17.64	2404	I 09QX47 mus musculus (mouse)
SP_human:012955	-	105.50	109.35	21.95	4377	I 012955 homo sapiens (human)
SP_human:014881	-	105.00	129.41	11.79	622	I 014881 homo sapiens (human)
SP_invertebrate:09A4U7	+	105.00	128.87	12.01	654	I 09A4U7 caulobacter crescentu
SP_invertebrate:061228	-	105.00	126.60	12.98	810	I 061228 tenobrio molitor (y
SP_bacteria:09ALM2	+	104.50	105.36	28.67	5588	I 09ALM2 saccharopolyspora sf
SP_plant:09PE83	+	104.00	127.24	14.47	669	I 09PE83 oryza sativa (rice).
SP_human:000508	+	104.00	118.05	19.81	1587	I 000508 homo sapiens (human)
SP_human:099552	-	104.00	109.43	26.62	3570	I 099552 homo sapiens (human)
SP_bacteria:09K7M6	+	103.50	133.02	12.67	364	I 09K7M6 bacillus halodurans
SP_bacteria:0918C7	+	103.50	101.19	37.68	7257	I 0918C7 polyangium cellulost
SP_bacteria:09ZFB9	+	103.00	116.70	23.65	1581	I 09ZFB9 rhizobium etli. glut
SP_fungi:096WV6	-	103.00	106.91	33.07	3971	I 096WV6 schizosaccharomyces
SP_human:09H773	+	102.50	135.33	13.34	257	I 09H773 homo sapiens (human)
SP_bacteria:09Z6E5	+	102.50	124.26	19.49	728	I 09Z6E5 listeria monocytogen
SP_bacteria:09BXA4	+	102.50	123.54	19.98	779	I 09BXA4 streptomyces aeruginos
SP_bacteria:09Z1W5	+	102.50	121.53	21.40	941	I 09Z1W5 streptomyces coelicol
SP_human:09Y2N2	-	102.50	118.14	24.03	1294	I 09Y2N2 homo sapiens (human)
SP_invertebrate:09NKS5	-	102.50	114.13	27.57	1866	I 09NKS5 leishmania major.
SP_invertebrate:09NK11	-	102.50	110.91	30.78	2554	I 09NK11 leishmania major.
SP_bacteria:09RAH4	+	102.50	99.80	45.03	4379	I 09RAH4 nostoc sp. gsv224. r
SP_bacteria:09K1Z7	+	102.50	99.80	45.03	7257	I 09K1Z7 polyangium cellulost
SP_bacteria:09S5M5	+	102.00	129.44	17.42	419	I 09S5M5 escherichia coli. udt
SP_bacteria:051415	+	102.00	128.23	20.83	684	I 051415 pseudomonas aeruginos
SP_vertebrate:073749	-	102.00	118.08	25.71	1219	I 073749 xenopus laevis (afri
SP_bacteria:09PC50	+	101.50	134.37	15.71	247	I 09PC50 streptomyces coelicol
SP_bacteria:09F617	+	101.50	127.96	19.56	451	I 09F617 synechocystis sp. (st
SP_human:014395	-	101.50	126.76	20.39	505	I 014395 homo sapiens (human)
SP_bacteria:098K15	+	101.50	126.55	20.53	515	I 098K15 rhizobium loti (mesor
SP_bacteria:093S58	-	101.50	126.38	21.37	575	I 093S58 myxococcus xanthus. u
SP_fungi:074851	-	101.50	118.18	27.34	1131	I 074851 schizosaccharomyces
SP_human:09P1P3	-	101.50	110.72	35.30	2281	I 09P1P3 homo sapiens (human)
SP_human:09Y618	-	101.50	110.54	35.53	2321	I 09Y618 homo sapiens (human)
SP_bacteria:098KX5	+	101.00	127.77	21.02	430	I 098KX5 rhizobium loti (mesor
SP_archaeap:09HRT5	+	101.00	124.98	23.13	559	I 09HRT5 halobacterium sp. (st
SP_bacteria:098DA7	+	101.00	122.68	25.02	694	I 098DA7 rhizobium loti (mesor
SP_plant:09Y8W8	+	101.00	120.03	27.41	894	I 09Y8W8 arabidopsis thaliana
SP_invertebrate:09N8G4	-	101.00	119.99	27.44	1249	I 09N8G4 trypanosoma brucei
SP_invertebrate:09VND0	+	101.00	109.96	38.69	2296	I 09VND0 drosophila melanog
SP_bacteria:09K1T5	+	100.50	130.12	20.70	323	I 09K1T5 vibrio cholerae. fruc
SP_bacteria:006470	+	100.50	127.72	22.48	405	I 006470 butyrivibrio cecapica
SP_bacteria:0985C9	+	100.50	120.05	29.23	833	I 0985C9 rhizobium loti (mesor
SP_bacteria:098PM7	+	100.50	119.40	29.89	885	I 098PM7 myxococcus xanthus. f
SP_plant:09HMT6	+	100.50	117.64	31.75	1045	I 09HMT6 oryza sativa (rice).
SP_human:09K9W9	-	100.50	116.01	33.56	1217	I 09K9W9 homo sapiens (human)
SP_bacteria:0918L1	+	100.50	115.44	34.23	1285	I 0918L1 shigella flexneri 2e
SP_human:09QNM47	-	100.50	109.15	42.46	2321	I 09QNM47 homo sapiens (human)
SP_invertebrate:023587	-	100.50	104.76	49.34	3307	I 023587 caenorhabditis ele
SP_bacteria:09EVN5	+	100.00	129.23	22.78	329	I 09EVN5 pseudomonas stutzeri

seq_name: sp_bacteria:044130

seq_documentation_block:

ID 044130 PRELIMINARY;

PRT; 449 AA.

AC 044130; 044130;

DT 01-NOV-1996 (TREMUREL. 01, Created)

DT 01-NOV-1996 (TREMUREL. 01, Last sequence update)

DT 01-NOV-1998 (TREMUREL. 08, Last annotation update)

DE 48 KDA OUTER MEMBRANE PROTEIN.

GN AOPA.

OS Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales; Pasteurellaceae;

RP SEQUENCE FROM N.A.
 RC STRAIN=M50;
 RA Ross B.C., Barr I., Patterson M., Agius C., Rothel L., Margetis M.,
 RT Hocking D., Webb E.;
 "Porphyromonas gingivalis polypeptides and nucleic acids."
 RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF144076; AAD3930.1;
 SO SQUENCE 451 AA: 49825 MW: 3EAC9E66FE5AD74 CR664;

alignment_scores:

Quality: 663.00 Length: 452
 Ratio: 2.225 Gaps: 7
 Percent Similarity: 65.929 Percent Identity: 34.735

alignment_block:

US-09-303-518D-125 x 09X6S0

Align seg 1/1 to: 09X6S0 from: 1 to: 451

```

1 ATGATTAAATCAAAAAGGCTTAACCTGCCATCGCGGAGACCGGA 50
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
4 ValIleThrlYsThrlYsGlyLeuAlaLeuAlaLeuGlyLeuProle 20
51 GCAAGCGCTTACGAGCGCGCCGCTTACCGAAGTC...GGTTGCTTG 97
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
20 uProGluMetLeuAlaGluProAlaGlnSerProThrlYrAlaValaP 37
98 GCGAGATATGCGGCTATGCGCCCTCGATGAAAGTCAAGAACGCGAT 147
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
37 roAspAspPheGluGlyValIleProGlyValThrlAlaArgProGly 53
148 GCGGTCAAAAAGCCCAAGTGGTGTGGAAGCAAAAAGAAATCGGGGCT 197
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
54 LysValAlaGlyAlaGlySerAlaLeuMetHisLysAlaTyProGluMe 70
198 GGTGTTTACGCGCGCTTACGCAAAATCCCGCATTCAGCTGGCG 247
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
70 LysPheThrSerProValSerGlyGluValIleAlaValAsnArgGly 87
248 AAAAGCGCTTACTGATGCTGATGATGCGGTGGAAGCAAGCAAGCA 297
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
87 LysArgGlyValLeuSerIleGlyValLysProAspGlyLeuAsnGlu 103
298 ATCGAG...TTGAAGCGTACGCAAGCTGCGGCTGCAACTTAAGCG 344
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
104 TyrGlnSerPheProValGlyAspProSerAla.....LeuSerAl 117
345 GCAAGAGTGGCGCGCAACGTATCCATTCGCTTGTGACTCGCTGC 394
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 acGluGlnIleLysGluLeuLeuSerSerGlyMetIlePheLeu 134
395 GCACCGCGCTGCTGAGCAAAATTCCTGCGCGTGCAGACCGCTTGGC 444
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
134 YsGlnAlaProIleAspIleValAlaThrProAspIleAlaProArgAsp 150
445 ATCTTGCTGATGCGATGAGACCAATCCGCTGCTGCGAGCCCTACGCT 494
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 IleTyrIleThrAlaAsnPheThrAlaProLeuAlaProAspPheAsp 167
495 CATATCAAAAGACCGCGCGGATTTCAACGCGGCTTGTGATTTGA 544
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 eIleValAlaGlyGluGlnAlaArgAlaLeuGlnThrAlaIleAspAlaLeu 184
545 GCGGTTTGACGAGCAAGCAAAATCGATGTTGTAAGCAGCTGCGAGAC 594
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
184 LysLeuThrThrGlyLysValAlaValAlaLysLeuLysProGlySerSer 200
595 GTGCGGCTGAAAATGCTGCCAATCGAAGCAACATGAATTTGCGCGGCC 644
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
201 LeuGlyLeuHisAsnAlaGluIleValGluValHis.....GlyTyr 214
645 GCATCTGCGCGGTTGAGTGAGCAAGCAATTCATTTCATCGAGCGGCTG 694

```

```

||||| ||||| ||||| ||||| ||||| ||||| |||||
214 OhisProAlaGlyAsnValAlaGlyValLeuIleAsnHisThrLysProIleAla 231
695 GCGGCAATAAACCGCTGTGACCATTCATTAATCAAGATGAATTAACCAT 744
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
231 snArgGlyGluThrValThrLysAlaThrAspLeuIleValIle 247
745 GCGCGTTGTTGCAACAGCGCGCTGCAACAGCGGCTGATTCCT 794
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
248 GlyArgPheLeuLeuThrGlyLysAlaAspPheThrArgMetIleAlaLe 264
795 AGTGCTTTCATGCAACCAACCGCGCTGCTGCTGCTGCTGCTGCTGCTG 844
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
264 ThrGlySerAspAlaAlaAlaHisGlyTyrValaArgIleMetProGly 281
845 CGAAGATATGCAAAATTCATCGCGCGCATTCGTT.....GACACAGAC 888
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
281 YsAsnValPheAlaSerPheProGlyArgLeuThrIleLysGlnSerHis 297
889 AACCGCGTATTCGCGTTCGATGATGAAAGCGCGGCTGATACACAGCGGC 938
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
298 GluArgValIleAspGlyAsnValLeuThrGlyLysLeuGlyGly 314
939 GCACGATATTTGGACGCTACCAACATCAATTCGCTTATGAAAG 988
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
314 sGluProPheLeuSerAlaArgCysAspGlnIleThrValIleProGlu 331
989 GCGCGAGC...AAGAGCTGTCGCGTGGTGGTGGCGCGCGCGCGCAAA 1035
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
331 LysAspAspAlaAspGluLeuPheGlyTyrPheAlaAlaProArgLeuAsp 347
1036 TACTCCATACCGGTACCAACCTCGCGCATTCCTG...AAAAACAAC 1082
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
348 TyrSerMetSerArgAlaTyrPheSerTrpLeuGlnGlyLysAsnArg 364
1083 CTTCAGATGATGCAACAGCGCGTGAATGCGTGTGATTCGCGCACCTGCT 1182
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
364 uTyraValLeuAspAlaArgIleLysGlyGluAlaGluAlaMetIleMet 381
1133 TTGTTACTTACGAGCGCGTGAATGCGTGTGATTCGCGCACCTGCT 1182
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
381 eAsnGluTyrAspArgValPheProMetAspIleTyrProGluTyrLeu 397
1183 TTGCGCGATTAATGTCGCGGATACGACAGCGCGCGCATTCGCTG 1232
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
398 LeuLysAlaIleIleAlaPheAspIleAspLysMetGluAspLeuGly 414
1233 CTTCGATGATGCAACAGCGCGTGAATGCGTGTGATTCGCGCACCTG 1282
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
414 eTyraValAlaProGluAspPheAlaThrCysGluPheValAspHis 431
1283 GCAATATGATACGCGCGCGCTGTCGCAAAAGTGTGGAACCATTCG 1332
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
431 eLysIleGluLeuGlnArgIleValaArgGlnGlyLeuAspMetLeuTyr 447
1333 AAGGAA 1338
   |||||
448 LysGlu 449

```

seq_name: sp_bacteriap:Q9WY86

seq_documentation_block:
 ID Q9WY86 PRELIMINARY; PRJ: 451 AA.
 AC Q9WY86;
 DT 01-NOV-1999 (TREMUREL. 12, Created)
 DT 01-NOV-1999 (TREMUREL. 12, Last sequence update)
 DE ELECTRON TRANSPORT COMPLEX PROTEIN, PUTATIVE.
 GN TM0244.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogales; Thermotoga.
 OX NCBI_TaxID-2336;
 RN [1]

SEQUENCE FROM N.A.
 RC STRAIN-MSB8 / DSM 3109;
 MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
 RA Hart D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 genome sequence of *Thermotoga maritima*.";
 RL Nature 399:323-329(1999).
 DR TIGR; TM0244; -
 DR EMBL; AE01708; AAD3335.1; -
 DR InterPro; IPR001450; 4Fe4S-ferredoxin.
 DR InterPro; IPR001949; Complex1_51K.
 DR Pfam; PF00037; fer4; 1.
 DR PROSITE; PS00198; 4Fe4S-FERREDOXIN; 2.
 DR Iron-sulfur; Complete proteome; 7B2D7272CFC6E418 CMC64;
 SQ SEQUENCE 451 AA; 49515 MW; 7B2D7272CFC6E418 CMC64;

alignment_scores:
 Quality: 140.00 Length: 454
 Ratio: 0.648 Gaps: 17
 Percent Similarity: 47.57 Percent Identity: 20.264

alignment_block:
 us-09-303-518d-125 x Q9WY86 ..

Align seg 1/1 to: Q9WY86 from: 1 to: 451

```

49 GAGCAGCCGTTTACGACGCGCCGATACCGAAGTCGGTGGTGG 98
   ::::: ::::: ::::: ::::: :::::
35 AspyrProIIeGluArgAlaProLeuProGlnLysValPheValPheIle 51
   ::::: ::::: ::::: ::::: :::::
99 CGAAGAAATGCGGTATGCGCCCGCATGAAGTCAAGAGAGCGATG 148
   ::::: ::::: ::::: ::::: :::::
51 userAmHisAlaGlyAsnProAlaLysProValSerProGlyLysP 68
   ::::: ::::: ::::: ::::: :::::
149 CCGTCAAAAAGGCCAAGTCTGTTTGAACAAAAAGAAATCCGGCGGTG 198
   ::::: ::::: ::::: ::::: :::::
68 LuValLysThrGlyValIleGlyGluProGlnGlyPheIleSerAla 84
   ::::: ::::: ::::: ::::: :::::
199 GGTGTACTGCGCGCGCTTCAGGAAATGCGCGCATTCACCGTGGCGA 248
   ::::: ::::: ::::: ::::: :::::
85 TyrLeuHisSerProValThrGlyArgValLeuGluIle..... 97
   ::::: ::::: ::::: ::::: :::::
249 AAGCGCGTACTT.....CAGTCACTGCTGA 274
   ::::: ::::: ::::: ::::: :::::
98 LysLysIleLeuHisProIleLeuGlyLysProIleGluAlaIleValI 114
   ::::: ::::: ::::: ::::: :::::
275 TTGCGCGTTGAGGACGACGCAATC.....GAGTTT 306
   ::::: ::::: ::::: ::::: :::::
114 LeGluArgThrSerAspAspGluTrpValHisIleGluThrGlyAspPhe 130
   ::::: ::::: ::::: ::::: :::::
307 GAACGCTAGCAGCTGAAGCGCTGGCAAACTTAAGCGCGAAGAGTGGC 356
   ::::: ::::: ::::: ::::: :::::
131 GluArgMetSerLysGluGluLeu.....GluIleI 142
   ::::: ::::: ::::: ::::: :::::
357 CCGCAACTGATCAATCCGGTTTGTGACTCGGCTGGCGACCCGTCGGT 406
   ::::: ::::: ::::: ::::: :::::
142 euLysLysAlaGlyIleValIleGluGlyGlyAlaMetPheProThrHis 159
   ::::: ::::: ::::: ::::: :::::
407 TCAGCAAAATTCCTGCGCTGCATGCGAGCGCTTCGCATTCGTCAT 456
   ::::: ::::: ::::: ::::: :::::
159 alLysLeuSerProProGluLysLysValAspThrLeuIleValAsn 175
   ::::: ::::: ::::: ::::: :::::
457 GCGATGAGCAACATCCGCTGGCTGCCAGACCTACGCTCATTAACAAGA 506
   ::::: ::::: ::::: ::::: :::::
176 GlyAlaGluCysGluProValLeuThrIleAspHisArgLeuMetLeuG 192

```

```

507 AGCGCGGAGATTCAAACGCGCGCTGTGTATGACCGCTTGGACG 556
   ::::: ::::: ::::: ::::: :::::
192 uArgAlaAspIleLeuGlnGlyIleLeuIleMetLeuLysVal.... 207
   ::::: ::::: ::::: ::::: :::::
557 AACGCAAAATTCATTTGTGAAGCA...GCTGGCGGACGCTCGCT 603
   ::::: ::::: ::::: ::::: :::::
208 .....LeuGlyValGlnLysAlaValAlaGlyValGluSerAsnLys 221
   ::::: ::::: ::::: ::::: :::::
604 GAAATGCGCCACATCGAATGCAATTCGCGCGCGCGCATTCCTGCG 653
   ::::: ::::: ::::: ::::: :::::
222 MetAspAlaThrHisLeuLysLysValPheLysGly...TyrProVa 237
   ::::: ::::: ::::: ::::: :::::
654 CGGTGTAGTGGCAGCGCATTCATTCATCGACGCGCTGGCGCAATA 703
   ::::: ::::: ::::: ::::: :::::
237 LasValAlaLeuLeuArgThrLysTyr.....ProGlnGlyAlaGlu 252
   ::::: ::::: ::::: ::::: :::::
704 AA...ACCGTGGAGCATC..... 720
   ::::: ::::: ::::: ::::: :::::
252 ysdInLeuIleTyrAlaIleThrGlyArgMetValProArgGlyGlu 268
   ::::: ::::: ::::: ::::: :::::
721 .....AATATCAAGATGTAATTCAT 743
   ::::: ::::: ::::: ::::: :::::
269 ProMetAspValGlyValValAlaGlnAsnValGlyThrCysValAlaVa 285
   ::::: ::::: ::::: ::::: :::::
744 TGCGCGTTGTGTCACACAGCGCGTGTGAACCGAGCGCGTATGGCC 793
   ::::: ::::: ::::: ::::: :::::
285 lLysGluAlaValAlaValAspLysProLeuValGluArgGlyMetThrV 302
   ::::: ::::: ::::: ::::: :::::
794 TAGGTGCTCTCAAGTCAACAAACCGCGCTGTGGTACCGCTTTGGGT 843
   ::::: ::::: ::::: ::::: :::::
302 alSerGlyAspAlaValAlaLysAsnGlnLysAsnLeuIleValArgIle 318
   ::::: ::::: ::::: ::::: :::::
844 GCGAAGTATGCAATTAATCT.....GCGGCGAATGTGTGACACAGA 887
   ::::: ::::: ::::: ::::: :::::
319 ThrProValLysAspValIleAspTyrCysGlyGlyIleAspLysnTh 335
   ::::: ::::: ::::: ::::: :::::
888 CAACGCGGATTCGCGGTGCGGTATGAACGCG...CGATTAACAGAG 934
   ::::: ::::: ::::: ::::: :::::
335 rGluArgValIleLeuGlyLysProMetMetGlyIleSerIleThr.... 350
   ::::: ::::: ::::: ::::: :::::
935 GCGGCGACGATTAATTTGGAGCTACACATCAATGATTCCTGATTCGA 984
   ::::: ::::: ::::: ::::: :::::
351 .....AsnLeuAspIleProValMet 357
   ::::: ::::: ::::: ::::: :::::
985 GAAGCGCGACGCAAGAGCTGTTCGCTGGCTGGCGCGAGCGGACAA 1034
   ::::: ::::: ::::: ::::: :::::
358 LysGlyThrSer.....PheLeuProLys..... 361
   ::::: ::::: ::::: ::::: :::::
1035 ATATCCATCAGCGGTACAAACCTCGGCGCATTTCTGAAAACAACTCT 1084
   ::::: ::::: ::::: ::::: :::::
362 ...GlyThrThrAla.....PheLeuProLys..... 370
   ::::: ::::: ::::: ::::: :::::
1085 TCAGTTCACACAGCGCTACAGCGCGGCGAGCGCGCATGTCGGCAT 1134
   ::::: ::::: ::::: ::::: :::::
371 .....SerArgProGlnLysProCysIleLeuArgCys 380
   ::::: ::::: ::::: ::::: :::::
1135 GGTACTACGAGCGGATGATCCCTTGATATCCGCCACCTCGCTTTT 1184
   ::::: ::::: ::::: ::::: :::::
381 SerGluCysValGlnValCysProMetAsnLeuGlnProTyrLeuLeu 397
   ::::: ::::: ::::: ::::: :::::
1185 GCGCGATTAAATCGTGGCGATACCGACGCGCGCAGGCAATGGGTGCT 1234
   ::::: ::::: ::::: ::::: :::::
397 rLeuLeuSerThrLysArgLysTyrAspGluAlaValAlaGlnAsnGlyLeu 414
   ::::: ::::: ::::: ::::: :::::
1235 TGGATTTGGAGAGAAACACCTCGCTTGCACCTTCGTCGCGCGGCG 1284
   ::::: ::::: ::::: ::::: :::::
414 eTAspCysIleLeu.....CysGlySerCysThrTyrThrCysProSer 428
   ::::: ::::: ::::: ::::: :::::
1285 AATATGCAATAC 1296
   ::::: ::::: ::::: ::::: :::::
429 LysIleGluHis 432
   ::::: ::::: ::::: ::::: :::::
seq_name: sp_fungi:004051

```



```

753 .....SerSerThrThrSerSerProSerAlaProValAlaVal 767
172 ACAGCAGCTGGCCCTTTTGAGGAGATGCTCCTTACATTCATTCAG 123
767 etherThrThr.....ThrSerProSerAlaLeuValValLeu 780
122 GGGCGCATACCGGATATTC.....TCGCCAAGCAA 91
781 SerSerThrSerThrSerProThrAspIleValTyrSerProSerTh 797
90 CGCGACTGGATATGCGCGG.....CCGTCGTAACGCTTCCTCCG 47
797 rpeAlaAlaIleSerSerGlyTyrThrProSerProSerAlaSerVal 814
46 GTCGCCCGCGGATGGCAGGTTAGACCTTTTGTATTTTA 6
814 IatSerSerThrSerSerSerSerProThrAspIleVal 827
seq_name: sp-fungi:Q9C105

```

```

seq_documentation_block:
ID 09C105 PRELIMINARY; PRT: 1236 AA.
AC 09C105:
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 19, Last annotation update)
DE PUTATIVE GLUCOMYLASE 1 (ALPHA-1,4-GLUCAN GLUCOSIDASE), EXTRACELLULAR
DE STARCH-DEGRADING ENZYME, BY SIMILARITY TO S. CEREVISIAE STAL, CONTAINS
DE CHITINASE FAMILY SIGNATURE.
GN SPAPL17.04C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxId=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H;
RA Wood V., Rajandream M.A., Barrell B.G., Seeger K., Harris D.;
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL590605; CAC36921.1;
DR InterPro: IPR001917; AminoTransf.-2.
DR InterPro: IPR001579; Chitinase-2.
DR Pfam: PF00192; Chitinase_2; 2.
DR PROSITE, PS00599; AA-TRANSFER CLASS 2; UNKNOWN 1.
DR PROSITE, PS00599; AA-TRANSFER CLASS 2; UNKNOWN 1.
SQ SEQUENCE 1236 AA, 123387 MW, 5A2D33A30B87CDD8 CRC64;

```

alignment_scores:

Quality:	132.00	Length:	471
Ratio:	0.569	Gaps:	17
Percent Similarity:	49.257	Percent Identity:	20.807

alignment_block:

US-09-303-518D-125/rev x Q9C105

Align seg 1/1 to: Q9C105 from: 1 to: 1236

```

1331 TCAATGGTTTCAGCACTTGGCAACAGCGCGGCTATTCGATTTGCC 1282
||| ||||| ||||| |||||
505 SerGluValSerSerThrLeu.....LeuSe 513
1281 CGGCGAGAGAGAGCTGCACAAAGCGAGGCTTCTTCGTCATTC... 1236
||||| : : : : : ||||| : : : : :
513 rGlySerSerAlaIleProSerThrSerSerThrProSerSerSer 530
1235 .....AAGCAACCAATGCTCGCGCGCTGCGGATG.....CCG 1200
: : : : : : : : : : : : : : : : : : : : :
530 IatSerSerProMetThrSerValLeuSerSerSerSerIlePro 546
1199 ACGATTAAATCGCG.....AAAGCAGGCTGGCGAGATATCAAGG 1156
||| : : : : : : : : : : : : : : : : : : :

```

```

547 ThrSerSerSerSerAspPheSerSerSerIleThrIleSerSer 563
1155 CATCAGCGGCTCGTACATCA.....ATCG 1130
||| : : : : : |||
563 yIleSerSerSerSerIleProSerThrPheSerSerValSerSerIle 580
1129 GCACCATGGCGCGCGCGCGGCTGACGCGCTGTTGAATTCGAAAGT 1080
: : : : : : : : : : : : : : : : : : : : :
580 euSerSerSerThrSerSerProSerSerThrSerSerIleSerSer 596
1079 TTTCTTTTCAGGAAATGGCCGAGGTTGTACCGTGAATGATTTGTC 1030
597 .....Se 597
1029 CGGCTCGCGCGGCAACCGACAGCGCTTTGCGGCTTCTTCA 980
||| : : : : : |||
597 rSerThrSerSerThrPheSerSerIleThrSerSerProSerSer 614
979 TAACGGAATCTGATTTGTGAGCGTCCCAATATGTCGCCGCCCTTGT 930
||| : : : : : |||
614 IeSerSer.....SerIleSerSerSerThr 623
929 GTATTCGCG.....CCGTCATACCGACCGGAATCAGCGGTT 889
: : : : : : : : : : : : : : : : : : : : :
624 IleLeuSerSerProThrProSerThrSerSerLeuMetIleSerSer 640
888 GTCGTGTCAACCAATTCGCCCGCAGTAAATTTGCAATCTTCGCACCA 839
||| : : : : : |||
640 rSerIleIleSerGlySerSerSerIleLeuSerSerIleSer... 655
838 AAACGTCAGCAAGAGCGCGGTTGTACTTGAGAACCCATGAGGCA 789
||| : : : : : |||
656 ..ThrIle.....ProIleSer 660
788 ATACGCGGCTCGGCTTCAGACGCGCTTGGCAACAAACGCGCAATGT 739
||| : : : : : |||
661 SerSerSerSerThrThrSerSerSerValIleProSerSerThrIle 677
738 AATTACATCTGATTAATGATGCTCCACACGTTTATTTCGCGCGGACG 689
: : : : : : : : : : : : : : : : : : : : :
677 uValSerSerSerSerSerIle.....ValSerSerProValIle 692
688 GTCGTGATGAATGAATGTCGTCACGCACTCAACCGCAGAGCGCGCG 639
: : : : : : : : : : : : : : : : : : : : :
692 IatSerSerSerSerProIleProSerSerSerSerIleValSerThr 708
638 CCGAATTCATGTTTCG.....ATGTTGGCAGC 610
: : : : : : : : : : : : : : : : : : : : :
709 TyrSerAlaSerLeuSerAlaSerIleThrIleSerSerLeuSerThrAl 725
609 ATTTTCAGACGCGACGTCGCGCGAGCGCTTACAAACATGATTTGTC 560
||| : : : : : |||
725 ameSerSerSerSerAlaIleProThrSerValIleSerSerThrLeu 742
559 GTTCGTCMAAGCGCTCAATACCAACAGCGCGGTTGAATCTCGCGG 510
: : : : : : : : : : : : : : : : : : : : :
742 IerThrAlaSerSerSerAsnThrLeuSerSerIleThrSerSerSer 758
509 GCTTCCTTG...AATATGACCGTAGGTCGCGACCGCGGATGTGTC 463
||| : : : : : |||
759 AlaIleValSerSerThrThrValSerAlaIleSerSerAlaLeuRose 775
462 CATCGCATTCAGCAG.....ATGTTGGCAGC 447
||| : : : : : |||
775 rAlaThrIleSerSerGlnSerGlnLeuThrAsnSerSerThrLeuAla 792
446 .....ATGCGCAAGCGCTCGCGCATCGACCGCAATTTGCTG 408
: : : : : : : : : : : : : : : : : : : : :
792 hrSerLeuThrLeuSerSerSerSerSerSerArgThrIleSerThr 808
407 AACGAGCGGTGCGCGACGAGTGCACAAACCGGATTCAGTTCGTCG 358
: : : : : : : : : : : : : : : : : : : : :
809 ThrAsnGluTyrAsnThrSerPheHisAlaProThrValSerSerThr 825

```

```

357 GCGCAGCTTCCTCCGCCCTTAAGTTGCCAGCGCTTCAGTGCCTGAAGGTT 308
      ::::|||||:
825 rSerSerSerThrThrSerLeuAlaAlaSerLysLysVal...Asn 841
      ::::|||||:
307 CAACCTGAT.....TCGTGTCCTCTCA 232
      |||||
841 eArnsSerLleThrSerLeuAsnLeuGluSerThrSerSerValThrSer 857
      :|||:|||||:|||||:
281 ACGGCAATCAAGACTGACTGAAGTACGCGCTTTTCGACAGGTGAATCGC 232
      |||||
858 ThrAlaThrThrAsp..... 863
231 GGGGATTTTCCCGAAGCGCGCGCACTAAACACACCGCCGATTCCTTT 182
      ::::|||||:
864 .SerValThrSerThrThrAlaLeuThrSerGlnGlyProSerSerSerV 880
      :|||:|||||:
181 TGTCTTCAACACAGCACTGGCCCTTTTGAGAGGAGATCGCCTTCCTGACT 132
      :|||:|||||:
880 aAlaSerSerSerLeuSerSerThrThrSerLeuSerThrSerLlePro 896
      :|||:|||||:
131 TTCATGAGGGGCGCATACGGCATATTCCTTCCGCAAGCAACCGACTTC 82
      :|||:|||||:
897 ValThrSerSerValAlaProAlaValThrSerThrGlySerGluThrSe 913
      :|||:|||||:
81 GGTAAAGCGCCGG 69
      :|||:|||||:
913 rSerValValGly 917

```

seq_name: sp_human:O60382

```
seq_documentation_block:
ID      060382      PRELIMINARY;      PRT;      1791 AA.
AC      060382:
```

DT	01-AUG-1998 (TIREMBLrel. 07, Created)
DT	01-AUG-1998 (TIREMBLrel. 07, Last sequence update)
DT	01-DEC-2001 (TIREMBLrel. 19, Last annotation update)
DE	KIAA0324 (FRAGMENT).
GN	KIAA0324.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX	NCBI_TaxId=9606;
LN	[1]
RP	SEQUENCE FROM N.A.
RA	Ricke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,
RA	Robinson D., Jones M., Buckingham J., Chasteen S., Thompson S.,
RA	Goodwin L., Bryant J., Tesner J., Meincke L., Longmire J., White S.,
RA	Ueng S., Tatum O., Campbell C., Pawcett J., Deaven L.,
RT	"Sequencing of Human Chromosome 16p13.3."
RL	Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RP	[2]
RP	SEQUENCE FROM N.A.
RA	Ricke D.O.;
RT	"Large Scale Sequence Analysis and Annotation with the Sequence
RT	Comparison Analysis (SCAN) System."
RL	Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR	EMBL: AC0004493; AAC08453.1; -
FT	NON_TER
SO	SEQUENCE

alignment_scores:

Quality:	131.00	Length:	411
Ratio:	0.697	Gaps:	14
Percent Similarity:	45.742	Percent Identity:	25.061

alignment_block:

US-09-303-518D-125 X 060382

Align seg 1/1 to: 060382 from: 1 to: 1791

12 CAAAAGGTCTAACCTGCCCATCGCGGCGACCGGAGCAAGCGTTT 61

[illegible]

2037 leuprPargLysArgSerArgSerArgSerProleuAlaIleArgArg 2053
1024 CACCGGACAAATACTGCATACGCGTCAA..... 1054
|||||.....
2053 gserArgSerArgThrProArgThrAlaArgGlyLysArgSerIleuthra 2070
|||||.....
1055CCCGCGCCATTCTCTGAAAACAACTTTCAAGTCAACACAG 1099
||| ||||| |||||.....
2070 rgSerProAlaIleArgArgArgSerAlaSerGlySerSerSerasp 2086
1100 CCGTCACGGGCGGCGACCGCCCATGTGCGGA 1132
|||
2087 ArgSerArgSerAlaThrProProAlaIleThrArg 2097

```
seq_documentation_block:
ID      09U036      PRELIMINARY;      PRT; 1275 AA.
AC      09U036;
DT      01-MAY-2000 (TREMblrel, 13, Created)
DT      01-MAY-2000 (TREMblrel, 13, Last sequence update)
DT      01-DEC-2001 (TREMblrel, 19, last annotation update)
DE      RNA BINDING PROTEIN (FRAGMENT).
DE
DE
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RX      [1]
RN
RP      SEQUENCE FROM N.A.
RA      Cheeki S., Umeki K., Sawada Y.;
RT      "Homo sapiens mRNA for RNA binding protein, partial cds.";
RL      Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
FR      EMBL; AB016091; BAA83717.1; -.
SQ
SQ
SEQUENCE 1275 AA; 136869 MW; 45C2B2P5F5E98A6F6 CRC64.
1
1
```

```
alignment_scores:
    Quality: 131.00
    Ratio: 0.697
    Percent Similarity: 45.742
    Percent Identity: 25.061
```

```
alignment_block:
US-09-303-518D-125 x Q9UQ36
.
```

Align seg 1/1 to: Q9UQ36 from: 1 to: 1275

```

12 CAAAAAGGCTAAACCTGCCATCGGCGGCGAGCGAGCAAGCCGTT 61
   :::::|||||  ::  |||  ::|||  ::  |||
267 ArgArgSerAlaSerSerProArgThLysThlThSerArgArgI 28:
   |||  |||  :::::|||||  |||  |||  |||  |||
62 ACGACGGCCC.....GGCCATTACCGATCGCGTT... 92
   |||  |||  |||  |||  |||  |||  |||  |||
283 yArgSerProSerProLysProArgLysLeuGlnArgSerArgSerArgS 300
   .....GCTTGGCGAAGAAATATGCGGTATCGGCC 122
300 eArgArgGlyLysThlArgThlThlArgArgArgAspArgSerGlySer 312
   ::  ::  |||||  |||  |||  ::
123 CTCGATGAAGTCAAGCAAGCGATCCGTCAAAAAGGCCAAGTCTGT 172
   :::::|||||  |||||  |||||  :::::  |||  ||
317 SerInSerThrSerArgArgArgGlnArgSerArgSerArgSerArgVa 335
   TTTGAAGCAAAAAAGAAATCCGGGCGGTCTTACTACGCCCGGCTTCAAGCC 222
   |  |||||  |||||  |||||  |||  ::  ::
333 lThrArgArgArgArgGlyGlySerGlyThrHis..... 344
223 AAAATCCCCCGGCTTACCGGTGCGCAAAAGCGGACTTTCAGTCAGTGT 272
   :::::|||||  ::  ::  :::::|||||  |||||
345 ..SerArgSerProAlaArgGlnGlnSerSerArgThSerSer..... 358
273 GATTGCGGTGAAGCGAAGCAAGAAATGAGTTGAACGCTACGCACCTG 322
   |||:::|||||:::|||||  |||  |||

```

```

359 .....ArgArghArghrAgelArghArghrProProthr.. 370
323 AAGCGTGGCAAACTTAAGCGCGAAGAAGTGGCGGCACACTTGATCCAA 372
371 .....SerArghLyArghSerArghrThSerProblArPro 384
373 TCCGGTTGTGGACTCGCGTGGCAACCGCTCCGTTTCACAAATTCCTGC 422
384 rPhyArghSerArghrArghAlaSerProAlArHlArghArghSerArgh 400
423 CGTCGATGCCGAAGCGGTTCCGCAATCTTCGTCAAGCATGAGACCAATC 472
401 SerArghThProleuIlSerArghrArgh..... 410
473 CGCTGGCTCCGACACCTTACGGTCATTTACAAGAAGCCGCCGAGATTTC 522
411 .....SerArghSerArgh..... 414
523 AAGCGGCGCTGTGGTATGTAGCGGTTTGACGAAACGCAAAATTCATGT 572
415 ...ThSerProValSer..... 419
573 TTGTAAAGCAGCTGGCGCGACACTGGCCGTCAAAATCTGCCAATCG 622
420 .....ArgArghrSer..... 426
623 AAGACATGAATTCGGCGCGCCCGCATCTCGCGTTTGAATGGCACGCAC 672
426 ghArSerArghrArghArghArghSerArghAlaSerProValSerAr 443
673 ATTCAATTTCATCGAGCC.....GGTCGGCGCGCAATPA 704
443 rghArghrSerArghrSerArghrProProValThArghrArghSerArgh 459
705 AACCTGTGGACACATCAATTATCAAGATGTAAATTAACACTTGGCCGTTGT 754
460 SerArghrThProThrThArghArghArghSerArghrThProProVa 476
755 TTGCACAGGCGCGTGTGAMACCGACCGCGGTGATTGCGCTAGTGGTCT 804
476 lThArghArghrSerArghrSerArghrThProProValThArghArghs 493
805 CAACTCAAAACACCGCCCTCTTGCGTACCGTTTGG..... 841
493 erArghSerArghThSerProIlThrArghArghrSerArghSerArghr 509
842 .....GTGGAAAGATATGCAAAATTAACGCGGGAATGGTGGACACA 885
510 SerProValThArghrArghrArghSerArghThSerProValThAr 526
886 GACACACGCGGTGATTCCGGTTCGGTATTGAACGGCGGATTAACACAAG 935
526 ghArghrSerArghrSerArghrThSerProValThArghrArghSerArghs 543
936 CGGCGACAG.....ATTATTGGGAGCGTACACACATCAAGATTCCGTTA 979
543 erArghrThProProAlAlAlArghArghArghSerArghrThProleu 559
980 .....TCGAAGAAGCGCGACAAAGACTTTCGCGTGGGTGGCGCGC 1023
560 LeuProArghLyArghrSerArghrSerArghrProleuAlAlArghArgh 576
1024 CAGCGCGAACAATACCTCAATCAGCGTTCAC..... 1054
576 gSerArghSerArghrThProArghThArghAlArghLyArghSerLeuThAr 593
1055 .....CCGTGCGCAATTTCTGAAAACAACAACTTCAATGATCAACAGC 1099
593 rghSerProProAlAlAlArghrArghrSerArghAlaSerghLySerSerArp 609
1100 CCGTCACAGCGCGGACCGCGGCATGGTGGCGGA 1132
610 ArgSerArghSerArghThProProAlArghArgh 620

```

seq_name: sp_human:015038

seq_documentation_block:
 ID: 015038 PRELIMINARY; PRT: 1783 AA.
 AC 015038;
 DT 01-JAN-1998 (TReMBLrel. 05, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)
 DE KIAA0324 PROTEIN (FRAGMENT).
 GN Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN.
 RX MEDLINE=97349984; PubMed=9205841;
 RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
 Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. VII.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro."
 RL DNA Res. 4:141-150(1997).
 DR EMBL; AB002322; BAA20782.2; .
 FT NON_TER 1
 SQ SEQUENCE 1783 AA; 190940 MW; 660302F6FD4179AB CRC64;

alignment_scores:
 Quality: 131.00 Length: 411
 Ratio: 0.697 Gaps: 14
 Percent Similarity: 45.742 Percent Identity: 25.061

alignment_block:
 US-09-303-518D-125 x 015038 ..

Align seg 1/1 to: 015038 from: 1 to: 1783

```

12 CAAAAAGGTCTAAACCTGCCATCGGGGACACCGGACGAAGCGCTT 61
   :::::::::::::::::::: :::: :::: ::::
775 ArgArgArgSerAlaSerProArgThrLysThrThrSerArgArg 791
   :::::::::::::::::::: :::: :::: ::::
62 ACGAGGGGCC.....GGCCATTACCGAGTCGGCT..... 92
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
791 yArgSerProSerProLysProArgGlyLeuGlnArgSerArgSer 808
   :::::::::::::::::::: :::: :::: ::::
93 .....GCTTGGCGAAGATATGCGCGTATGCGGCC 122
   :::::::::::::::::::: :::: :::: ::::
808 erArgArgGluLysThrArgThrThrArgArgArgAspArgSerGly 824
   :::::::::::::::::::: :::: :::: ::::
123 CTCGATGAAGTCAAGAGAGCGGTGCGCAAAAAAGGCCAGTCTCTG 172
   :::::::::::::::::::: :::: :::: ::::
825 SerGlnSerThrSerArgArgGlnArgSerArgSerArgSerArg 841
   :::::::::::::::::::: :::: :::: ::::
173 TTGAAGACAAAAAGATCCGGCGTGTGTTACTGCGCGGCTTCAGGC 222
   | ||| :::::::::::::::::::: :::: ::::
841 lThrArgArgArgArgGlyGlySerGlyTyrHis..... 852
   :::::::::::::::::::: :::: :::: ::::
223 AAAATCGCCGCGATTACCGTGGCAAAAGCGCTACTTCAGTCAGCGT 272
   :::: :::: :::: :::: :::: ::::
853 ..SerArgSerProAlaArgGlnLysSerArgThrSer..... 866
   :::: :::: :::: :::: :::: ::::
273 GATTCCCGTTGAGGACGAGCAATGAGTTGAAAGCTACGACCTG 322
   ||| :::::::::::::::::::: :::: ::::
867 .....ArgArgArgArgGlyArgSerArgThrProProthr... 878
   :::::::::::::::::::: :::: ::::
323 AAGCGCTGGCAAACTTAAGCGCGAGAGAGTCCCGCAACCTGATCCAA 372
   :: :::: :::: :::: :::: ::::
879 .....SerArgLysArgSerArgSerArgThrSerProAlaProt 892
   :::::::::::::::::::: :::: ::::
373 TCCGGTTTGACCTGCGCTGCGACCGCTCCGTTACGCAAAATTCCTGC 422

```

```

892 rPlysArgSerArgSerArgAlaSerProAlaThrHisArgSerArg 908
   :::: :::: :::: :::: :::: ::::
423 CGTCGATGCCGAGCGGCTTGCATCTTGCATTCGATGACACCAATC 472
   :::: :::: :::: :::: ::::
909 SerArgThrProLeuIleSerArgArg..... 918
   :::::::::::::::::::: :::: ::::
473 CGCTGCTGCCGACCCGACGTCATTATCAAGACCGCGGAGATTTC 522
   :::: :::: :::: :::: ::::
919 .....SerArgSerArg..... 922
   :::::::::::::::::::: :::: ::::
523 AAACGCGGCTTGTGATTTGAGCGGTTTGACGAGCAAAATTCATGT 572
   ||| :::::::::::::::::::: :::: ::::
923 ..ThrSerProAlaSer..... 927
   :::::::::::::::::::: :::: ::::
573 TTGTAAAGCAGCTGCGCGACGTCGCGTGTGAAAATGCTGCAACATCG 622
   ||| :::::::::::::::::::: :::: ::::
928 .....ArgArgArgSer.....ArgSer 934
   :::::::::::::::::::: :::: ::::
623 AAACACATGAATTCGCGCGCGCATCCTCGGCTTGTGAGTCACGAC 672
   :::: :::: :::: :::: ::::
934 gThrSerValThrArgArgArgSerArgAlaSerProAlaSer 951
   :::::::::::::::::::: :::: ::::
673 ATTCAATTCATGAGCC.....GGTCGCGCGGAATTA 704
   :::: :::: :::: :::: ::::
951 rArgArgSerArgSerArgThrProProValThrArgArgSerArg 967
   :::: :::: :::: :::: ::::
705 AACCGGTGACCATCAATATTCAGATGATTAATCAATTCGCGGTTGT 754
   ||| :::::::::::::::::::: :::: ::::
968 SerArgThrProThrThrArgArgArgSerArgSerArgThrPro 984
   :::: :::: :::: :::: ::::
755 TTGCAACAGCGCGCTCTGAACACCGGCGGATTCGCGGTTCT 804
   :::: :::: :::: :::: ::::
984 lThrArgArgArgSerArgSerArgThrProProValThrArgArg 1001
   :::: :::: :::: :::: ::::
805 CAAGTCACAAACCGCGCTTGTGCTACCGCTTTG..... 841
   :::: :::: :::: :::: ::::
1001 erArgSerArgThrSerProIleThrArgArgArgSerArgSerArg 1017
   :::: :::: :::: :::: ::::
842 .....GTGCGAAAGTATCGCAATTTACGCGGAGATTTGGTGACCA 885
   :::: :::: :::: :::: ::::
1018 SerProValThrArgArgArgSerArgSerArgThrProAlaThr 1034
   :::: :::: :::: :::: ::::
886 GACACACCGCGTATTCGCTTCGATTGAAAGCGCGGATTCACAAAG 935
   :::: :::: :::: :::: ::::
1034 gArgArgSerArgSerArgThrSerProValThrArgArgSerArg 1051
   :::: :::: :::: :::: ::::
936 CGCGCAGC.....ATTATTGGAGCGCTACCAATTCAGATTCCGTTA 979
   :::: :::: :::: :::: ::::
1051 erArgThrProAlaIleArgArgArgSerArgSerArgThrProLeu 1067
   :::: :::: :::: :::: ::::
980 .....TCGAAGAAGCGCGCAGCAAGAGCTGTGCGTGGCTTCGCGC 1023
   :::: :::: :::: :::: ::::
1068 LeuProArgLysArgSerArgSerArgSerProLeuAlaIleArgArg 1084
   :::: :::: :::: :::: ::::
1024 CAGCCGCAAAATCTCATCAGCGGTACA..... 1054
   :::: :::: :::: :::: ::::
1084 gSerArgSerArgThrProArgThrAlaArgGlyLysArgSerLeuTha 1101
   :::: :::: :::: :::: ::::
1055 .....CCCTGCGCATTTCTTGAAAAACAACATCTTCAGTTCAACAG 1099
   :::: :::: :::: :::: ::::
1101 rGserProProAlaIleArgArgArgSerArgSerArgSerArg 1117
   :::: :::: :::: :::: ::::
1100 CCGTCAACGCGCGGACGCGCCATGTGCGCA 1132
   ||| :::::::::::::::::::: :::: ::::
1118 ArgSerArgSerArgAlaThrProProAlaThrArg 1128
   :::::::::::::::::::: :::: ::::

```

seq_name: sp_fungi:094317

seq_documentation_block:
 ID 094317 PRELIMINARY; PRT: 534 AA.
 AC 094317;
 DT 01-MAY-1999 (TReMBLrel. 10, Created)

DT	01-MAY-1999 (TReMBLrel. 10, last sequence update)
DT	01-MAR-2001 (TReMBLrel. 16, last annotation update)
DE	HYPOHETHELICAL 53.4 KDA SERINE-RICH PROTEIN C215.13 IN CHROMOSOME II
DE	PRECURSOR.
GN	SPBC215..13.
OS	Schizosaccharomyces pombe (fission yeast).
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;
OC	Schizosaccharomyces.
OX	NCBI_TaxID=4896;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=972;
RA	Lyne M.; Rajandream M.A.; Barrell B.G.; Rieger M.;
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC	-1- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC	-1- SIMILARITY: TO YEAST STAL.
DR	EMBL: AL03534; CMA2127.1; -.
CC	Hypothetical protein; Glycoprotein; Signal.
KW	SIGNAL
FT	1
FT	14
FT	CHAIN 15 534
FT	CARBOHYD 31 31
FT	CARBOHYD 426 426
FT	534 AA; 53361 MW; D620C15C08A6988E4 CRC64;
SO	SEQUENCE
	POTENTIAL.
	HYPOHETHELICAL SERINE-RICH PROTEIN C215.13
	N-LINKED (GICNAC. . .) (POTENTIAL).
	N-LINKED (GICNAC. . .) (POTENTIAL).
	D620C15C08A6988E4 CRC64;

```

alignment_scores:
  Quality: 126.00
  Ratio: 0.603
  Percent Similarity: 50.361
  Length: 415
  Gaps: 15
  Percent Identity: 20.241

```

```
alignment_block:
```

US-09-303-518D-125/rev x 094317

Align seg 1/1 to: 094317 from: 1 to: 534

```

1280 GGGCAGCAGCAGCTGCACAAAGCCAGGCTCTTCTGTCACATTCACGA 1231
      ||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
67 G1gYlYhPrOthYrThYrYg1YrYlAlaThPrOthSerSerSerg1 83
      ||| : : : : : : : : : : : : : : : : : : : : : : : :
1230 ACCCAATGCGTCCGCCGCTGCGGTATCCGCCAGATTAATCGGCAGAA 1181
      ||| : : : : : : : : : : : : : : : : : : : : : : : :
83 uPserIllePheserGlusSerIlaThPrOserGluthrIshnSerTyrs 100
      ||| : : : : : : : : : : : : : : : : : : : : : : : :
1180 GCAGGGTGGGCAGATATCCAAAGGCATCACCGCCTGTAGTACCATC 1131
      ||| : : : : : : : : : : : : : : : : : : : : : : : :
100 eSer : : : : : : : : : : : : : : : : : : : : : : : :
      ||| : : : : : : : : : : : : : : : : : : : : : : : :
1130 GGCACCATGGCGCGCTGCCGCCGCTTGACGCTGTGTGAATTCGAAG 1081
      : : : : : ||| ||| ||| : : : : : : : : : : : : : : : :
104 SerSerTy : : : : : SerAspProAlaThrSerGlutSerIleuProSerSerTh 118
      : : : : : : : : : : : : : : : : : : : : : : : :
1080 TTTGTTTTTCAGGAATGGCCGAGGGGTGTACGCGTATGAGATTTGT 1031
      : : : : : ||| ||| : : : : : : : : : : : : : : : :
118 rSerPhePheSerPro : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : :
1030 CCGGCTGGGGCAACCCAGCCAGACGCTCTTGCTGCGCGCTTCG 981
      : : : : : : : : : : : : : : : : : : : : : : : :
130 ro : : : : : SerSerThnGlusSerSerIleuAspProSerSer 143
      : : : : : : : : : : : : : : : : : : : : : : : :
980 ATACGGAAATCTGATTTGTGTAGCGTCCCAAAATATATCTGCGCGCTTG 931
      : : : : : : : : : : : : : : : : : : : : : : : :
144 ValSerSer : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : :
930 TGTATGTCGCGCTTCATATACGAACCGGAATACCGGTTCTGTGT 881
      : : : ||| ||| : : : ||| : : : ||| : : : ||| : : : |||
147 AlilIleuProSerSerThSerValGlVal : : : : : SerIleS 160
      : : : : : : : : : : : : : : : : : : : : : : : :
880 CAACCAATTCGCCCGCAGATTAATTTGCGATCTTGCACCCAAACAGGTA 831
      ||| : : : : : ||| : : : : : : : : : : : : : : : : :
160 eSerSerSer : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : :
830 CGCAAGAGCGCGGTTTGTGTACTGAGAACACACTTATGGGGAATACGCG 781

```

```

164 .....leuSerSerSerAspProLeuThrSerSerThrph 175
780 CTCGGTGTTCAGACGGCGCTGTGGAAACAAACGCCCAATGTAATTACAT 731
111 :::: ::::::::::::::::::::|||
175 eSerSerLeuSerSerSerThrSerSerSerGlnProSerVal.....188
730 CTTGATTAATGATGATGTCACACAGGTTTATTCGGCCGACGGCGTCATG 681
190 .....SerThrSerSerSer.....SerThrSerSerSer 195
680 AAATGATGTCGTGGCAGCTCAACACCGGACGAGATGCGGCGCCGCAATTC 631
196 .....ThrPheSerSerAlaIaIaProThrSerTh 205
630 ATGCTTTTCAGATGTGGCAGCATTTTCA.....GACGACGCTGCGCG 587
::: ||| ::::::::::: ||| ::::::::::: ||| ::::
205 rSerSerSerThyLeuSerSerSerSerSerValIaIaSerSerSerSerP 222
586 CAGCTCCCTTACAAACATGATTTTCGCTTCGCTCAACAGGCTCAATACC 537
||::: ::::: ::::: ||| ||| :::::
222 rSerSerSerSerSerSerThrLeuThrSerSerSerLeuSerThrSer 238
536 AACAGCGCGCGTTTGAAATCTCGCGCGGCTTTTATGATATGACCGTAG 487
::: ||| ::::::::::: :::::
239 SerIleProSerThrSerSerSerSer..... 248
486 GTCCGACGACGACGGATGTGTCATCGCATTCAGACGAAGATGCGACAG 437
||::: ::::::::::: ||| ||| ::::: ||:::
249 .SerThrSerSerSerLeuSerSerSerSerSerSerSerThrIaIaSerS 265
436 GCTGGCATTCAGACGCGACGAGATTTTGTCTGAACGACGAGGTCGCGACGCA 387
::: ||| ::::::::::: ||::: :::::
265 eSerSerSerSerSerSerIleIaIaSerSerSer.....SerSerSer 279
386 GTCACAAACCGGATGTGATCAGGTGCGCGCACTTCTTCGCGCGTTAA 337
::: ||| ::::::::::: ||:::
280 SerSerSerPro.....ThrSerThrSerSerThrIleSe 291
336 GTTTCGACGCGCTTCAGTGCCTGAGCGTTCGTAACACGATTCGTCGTGC 287
::: ||| ::::::::::: ||:::
291 rSerSerSerSerSerSerSerSerProThrSerThrSerThrIleS 308
286 CTTCAACGGGCATACACGACTCAGTACGACTACGCCCTTTTGGCCACGGTGA 237
||::: ::::: ||:::
308 eSerSerSerSerSer.....SerSerSerPheSerSerThrIleu 321
236 ATCCGCGCGATTTTTCCTGGAAGCCGCGGACATTAACACACGACCGCGATT 187
::: ::: ::::::::::: ||:::
322 SerSerSerSerSerSerSerSerSerSerPheSerSerPro..... 336
186 CTTTTCCTTCCTTCAACACGCACTTGCGCTTTTTCAGCGGACATGCGCTTCT 137
||| ||| ::::::::::: ||| ::::::::::: |||
337 .....ThrSerSerSerThrIleIaIaSerSerSerSerSerProSerS 352
136 TGACTTCATCGAGGGCGCATACCGCATATCTTCCGCAACGACGAGCG 87
::: ||| ::::::::::: ||| :::::
352 eSerSerPheSerSerSerThrThrSerSerSerIysSerSerSerSerPhe 368
86 ACTTGCGTATGCGCGCGCGCTGTAAACGCGCTTCTCGCGCTCG 42
::: ||| ::::::::::: ||| ||| :::::
369 SerSerThrValaSerSerSerSerThrSerSerSerSerThrIleu 383
seq_name: sp_bacteriap:Q98DE3
seq_documentation_block:
ID Q98DE3 PRELIMINARY; PRT; 646 AA.
AC Q98DE3
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE 2-DEHDRO-3-DEOXYGALACTONATE KINASE.
EN MER4742.

```

OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAF30309;
 RA MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 Takeuchi C., Yamada M., Tabata S.;
 RA "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti."
 RT DNA Res. 7:331-338(2000).
 RL EMBL: AP003004; BAB51328.1;
 DR Kinase; Complete proteome.
 KW SEQUENCE 646 AA; 68943 MW; 6AEAF6471BB6FB CRC64;

alignment_scores:
 Quality: 120.50 Length: 436
 Ratio: 0.739 Gaps: 23
 Percent Similarity: 37.385 Percent Identity: 23.853

alignment_block:
 US-09-303-518D-125 x Q98DE3 ..

Align seg 1/1 to: Q98DE3 from: 1 to: 646

```

26 ACCTGCCCATCGCGGCAAGCCGAGCAAGCCGTTACGACGCGCGGCC 75
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
23 Thrcysvalthrargmetargalalenihsalatyrsalalalargpr 39
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
76 ATTAACCAAGTCGCGT 110
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
39 oleugllyargalalyshpralavalallelyrilleuargmetilep 56
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
111 CGGTATCGCGCCCTCGATGAAGTCAAGAGGCGCGCTGTTACTGCG 160
   ||::  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
56 ro.phearglysserspargrpnearglyser..... 66
161 GCCAAGTCGTGTTGAAGACAAAAGATCCGGCGGTGTTACTGCG 210
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
67 ..... 68
211 CGGCTTCAGGCAAAATCGCCGATTCACCGTGGCGAAAAGCGCGTACT 260
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
68 lagly..... 74
261 TCAGTCAGTCGTGATTCGCGTGAAGCAAGCAAGCAAGTGAAGTGAAC 310
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
75 Alaileglylphrecysargprrogllyarg..... 85
311 GCTACGCACTGAAGCGCTGCAAACTTAAGCGCGCAAGAGTGGCGCGC 360
   |||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
86 .....leuglyalaasprhsargargargrpheargspa 97
361 AACSTGATCAATCGCGTGTGTCGACGTGCGCGACCGCGTGGTGGTCA 410
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
97 rgargglyalaasprarglyval...cysala.....Proglycysgin 110
411 CAAATTCCTCGCGCTCGATCGGCAAGCGCTGCGCATTCGTCATGCGA 460
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
111 .....Cysargleuhsarghsargly..... 118
461 TGGACCAACATCGCGTGGTGGCGACCGTACGCGCATTAATCAAGAGCC 510
   |||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
119 .glyserghisalaalacysargprohis.....A 129
511 GCGGAGATTCAAACGCGCGCTGTGTTGATGAGCGCTTGAACGAGC 560
   ||::|||::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||

```

```

129 rggluglylle.....Argproalalalaleupproglinspargrser 143
561 .....CAAAATTCATGTTGTA 577
144 Alaarghsargglyalaalargleuglyglyleuglygly..... 158
578 AGCGAGTGGCGGCGAGACGTGCGTGAATGCTGCAACATCAAGAGCA 627
   ||::|||::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
159 .glyalatprargcysghisglialaasprglnincysargleuglyarg 175
628 CATGAATTCGCGCG..... 641
175 laproargargargglyargasprglyvalleuglyglninprogly 191
642 .....CGCGATCGCGCTGTGAGTGGCGCGCATTCATTTCA 682
192 asprglnproglinalalaphelenuhsargalaglyglyleuglyarg 208
683 TCGACCGCGTGGCGCGCAATTAACCGTGTGGACCATTCATTCAGAGT 732
   ||::|||::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
208 sglualaglyargarg.....Argleuhsprhslnleu..... 219
733 GTAATTCATTCGCGCGTGTGTTGCAACAGCGCGTCAACACGAGCG 782
   |||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
220 .....Hisleuhsprleuhsprhslnproserarghis..... 231
783 CGTGAATGCGCTAGGTGTTCTCAAGTCAACAAACCGCGCTTGGCGTA 832
231 ..... 231
833 CGGTTTGGGTGCGAAATGTCGCAATTTACTGCGCGCGATTTGTTGAC 882
232 .....Alaglyleuhsargarglnleuglyln 241
883 ACAGACAAACG.....CGTGAATTCGCGTGGTATGGA 917
241 isproargprohinsglnlialatargdrglnalalargalaglyprohis 257
918 CGGCGCGATTCACAAAGCGCGCGAGATTA.....TT 949
258 proarggln...Argargargalalargleuglyasprhsarglnalal 273
950 TGGAGCGTACCAACATGATTCGTTATGAGAAAGCGCGCAAA 999
273 aglythleu.....glyasprargseralaserprocsarg 288
1000 GAGCTGTGCGGTGCGTGGTGGCGCGAGCGGCAAAATTCATCAACGCG 1049
288 lmlavalhinsglnllyglnhsalalaprogly..... 298
1050 TACAACCTCGCGCATTTCTGAAAACAAACCTTCAGTTCAACAGAG 1099
298 ..... 298
1100 CCGTCAACGCGCGGCGAGCGCGCATGTCGCGTGTGTTACTTACAGCGC 1149
299 .....Arghsarghsargly.....Alav 306
1150 GTGATGCGCTGTGATTCGCGCGCGCGCGCGTGTGCGGATTTAAGCT 1199
306 alproglyleuglycysleuhsalalalalalalalalalalalal 322
1200 CGGCGA 1205
323 Argarg 324

```

seq_name: sp_human:O9H9M1

seq_documentation_block:

ID O9H9M1 PRELIMINARY;

PRT: 639 AA.

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

01-DEC2001 (TrEMBLrel. 19, last annotation update)
DE NTRYEM002189 PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eulhelia; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxId=9606;
[1]
RA RM SEQUENCE FROM N.A.
RP Isogal T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Maetsuma T., Hosori T., Kaku Y., Kodaira H., Kondo H., Sugawara
RA Takahashi M., Chiba Y., Ishida S., Murakami K., Ono Y., Takiguchi
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
RA Minomura K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK002123; BAB14204.1; -
SQ SEQUENCE 639 AA; 62709 MW; 5A199FA67808F0DE CRC64;

```
alignment_scores:
  Quality: 119.50
  Ratio: 0.504
  Percent Similarity: 48.765
  Length: 486
  Gaps: 22
  Percent Identity: 22.634
```

```
alignment_block:
```

US-09-303-518D-125/rev x Q9H9M1

Align seg 1/1 to: Q9H9M1 from: 1 to: 639

1304	AGCGGGCCGATTTCGTATTTCGCCCGGCGACAGCAGCTTCACAAAGCGAG	1255
95	SerialProthetInSerLeuProthetProthetThSer.....	106
1254	GTCCTTCCTGCATATTCACAGCAACCAATGCCTCGCGCTGTGG...	1209
107	SerialAlaSerThSerAsnProAsnSerAlaSerLeuSerSery	123
1208GATCCGCCGACGATTAAATCCGGCGAA	1182
123	AlphealagLeuProLeuProLeuProThProThSerLeuLeuSer	139
1181	AGCAGCGTGGCGAGATATTCACAGGCATACGCGCTGTAGTA...CC	1135
140	AsnProThProValIleAlaIagLysSerThProSerValAlaGlyPr	156
1134	AATCGGCACCATGCGCGCGTCCCGCGCTTACAGCCT.....	1098
156	OleuGly.....ValAsnSerProLeuLeuSerAlaLeuLysGlyPhel	171
1098	1098
171	euthrSerAsnSphThraSLeuLLeaSnSerSerAlaLeuSerSerAla	187
1097GTGTGAACCTTGAAGATTGTGTTTTCACGAATAATGCC	1060
188	ValThSerGlyLeuAlaSerLeuSerSerLeuThraLeuGlnAsn....	202
1059	GAGGTTGTACCGCGATAGAGTATTTCGCGGCTGGCGGCGCAACCAGC	1010
203SerAspSerSerAlaSerLap	210
1009	CGAACAGCTCTTGTGTCGGCCCTTCTTCGATACGGAATCTGATTGTGG	960
210	roAsn.....LysSytTyraLapProSerAlaIlePro.....	220
959	TAGCGTCCCAATAATCTCGGCGCCTTGATATCGCGCGCTTCATATC	910
221	ThProGlnargThSerThPro...GlyLeuAlaLeuPhProcl	235
909	CGAACCGGAA.....ATACAGCGGTTGTCGTGTCA	878

253 yProProleuProValAlaAsnSerThrSerThrProleuThrLeuPro 252
877 CCAATGGCC... GCACTATTGGCCACTACTTTGCGACCCAA 837
252 aGInSerProleuAlaThrAlaLaserSerThrSerAlaProVal 268
836 ACGGTACGCAAGAGCGCGGTTTGTACTTGTAGAACACCTTGGGGGAT 787
269 SerGylSerSerAlaSerLeuAlaGlyProHisProoLynr 285
786 CAGCGGTGGGTTCACAGCGGCTGTGCAACAAAGCCCAATGATA 737
285 rAsPLeuHisLaserSerThrProAlaAlaThrLeuPro..... 299
736 TTACATCTGTAATTAATGATGTGTCACACGGGTTTATTTCGGCGGACGGC 687
300ValMetLLeuThrGluProThrSerProThrPro 311
686 TCGATGAATGAATGTGCGTCCGCACTGAA... CCGGCAAGATCGCGGC 640
312 Ser.....AlaPheLysGlyProSerHisSerLys 322
639 GCGCAAT.....TCATGTGTTGCAATGTTGGCAGCATTTTCAG 602
322 nProSerHisGlyThrLeuLysSerGlyThrLeuGlyAlaGlyAlaTyr 339
601 ACGGCACTGTGCGGCACTGCTGCTTACAAATGATGATTTGGCTGGGC 552
339 hSerThrSerValProLLeuSerLeuSerAlaCysLeuAsnProAlaLeu 355
551 AAACGGCTCAATACCAACAGAGCGGGTTGAA... TCCTGGCGGCTTC 505
356 SerGlyLeuSerSerSerThrProLeuAsnGlySerAsnProLeuSe 372
504 TTTGATTAATGACCGTAAAGGCTGGGACGCAACGCGATGTGTCATCGCAT 455
372 rSerLLeuSerLeuProProHisGlySerSerThrProLeuAlaProVal 389
454 TACAGCAATGTGCGCAACGGCTGGGATGAGCGGCAATTTGGTGAAC 405
389 herThrAlaLeuProSerPheThrSerLeuThrAsnAsnProleuThr 405
404 GGAGCG.....GTGCGAGCGCACTGCACAAACCGGATTTGGATC... 366
406 GlyAsnProSerLeuAsnProSerValSerLeuProoLysSerLeuAl 422
365AGTTGGCGGCACTTCTGCGCGCTTAAGTTGGCAGCGCTT 323
422 aThrSerSerThrAlaAlaThrSerThrSerLeuProHisProSerSer 439
322 CAGGTGCTAGCGTTTCAACATCGATTTGCGTGGCGCTTCAACGGCAATC 273
439 hrrAlaAlaValLeuSerGlyLeuSerAlaSerAlaProValSerAlaAla 455
272 ACGACTAGCTAACTAGCGGCTTTTGGCCACGGTAATCGCGGCGATTTT 223
456 Pro.....PheProLeuAsnLeuSerThrAlaVala 465
222 GCGTGAAGCGGCGAGTAAACACACAGCGCGCGGATCTCTTTGTCTTCAA 173
465 lProSerLeuPheSerVal.....ThrGInGlyProLeuSerSerSerA 480
172 ACAGACTTGGCCTTTTTCACGCGATCG.....CCTTCTGTGACTTTC 129
480 snLeuSerTyrProGlyPheSerValSerAsnThrProSerValThrPro 496
128 ATGAGAGGGGCGCATACCGGCGATTTCTGGCCACAGACGCGACTTCGGT 79
497 AlaleuProSerPheProoLysLeuGInAlaProSerThrValAlaAlaVala 513
78 AATGGCCGGGCG...TCGTAAACGGCTGTGCTCGGTGCGCGGCGATTC 32
513 lThrProleuProValAlaAlaThrAlaProSerProAlaProValLeuP 530

31 GCAGCTT 24
111
530 roglyphe 532

seq_name: sp_invertebrate:Q9VXM5

seq_documentation_block:

ID Q9VXM5 PRELIMINARY; PRT: 2406 AA.

DT 01-MAY-2000 (TRIMBLrel. 13, Created)

DT 01-MAY-2000 (TRIMBLrel. 13, Last sequence update)

DT 01-JUN-2000 (TRIMBLrel. 14, Last annotation update)

DE CG9056 PROTEIN.

GN CG9056

OS Drosophila melanogaster (Fruit fly)

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RA MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abil J.F., Agbayani A., An H.-J., Andrews Pfanhock C., Baldwin D.,

RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,

RA Boriva D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,

RA Burris K.C., Busan D.A., Butler H., Cadeu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,

RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Haris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., Mocherson D.,

RA Merklov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,

RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reineert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Strong R., Sun E.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Wang A.,

RA Svyrkas R., Teator C., Turner R., Venter E., Wang A., Wang X.,

RA Wang Z.-T., Wasserman D.A., Weinstock G.M., Weisenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong G.M., Zhou X., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

DR EMBL: AE003500; AAF48533.1;

DR FlyBase: FBgn0030702; CG9056.

SO SEQUENCE 2406 AA; 248803 MW; 740EAD78D64F76C CRC64;

alignment_scores:

Quality: 119.50

Ratio: 0.511

Percent Similarity: 47.561

Length: 492

Gaps: 27

Percent Identity: 21.951

alignment_block:

US-09-303-518d-125/rev x Q9VXM5

Align seg 1/1 to: Q9VXM5 from: 1 to: 2406

1340 CCTTCCTTCATGATGTTTCCAGCACTTTCGCAACGCGCGCATTC 1291
1278 ProValLeuProValAlaThProAsnLeuSerAsnLeuProThrGl 1294
1290 GTATTGCCCGGCGAGACGAGAGCTGCACAAAGGAGGCTCTTCGCA 1241
1294 nhisArgSerSerAspSerArg...AsnSerArgGlnSerProAlaSer 1310
1240 ATTCACACCAACCAATGCTGCGCGCTGCGATG...CCGAGC 1197
1310 euLysSerThProSerAsnIleGlyLeuAsnValSerMetAlaProThr 1326
1196 ATTAATG...CGCAAAAGAGGCTGGCGAGATTCACAAAGGCA 1153
1327 LeuArgSerIleThProLeuAsnAsnSerSerAlaIleSerSerGlyAl 1343
1152 CACGGCTGCTGAATGACCAATCGGACCATGCGCGCGCGCTTGA 1103
1343 aserGln...ProValSerValValProSerAlaAsnSer 1357
1102 CGGCTGCTGTAAGTGAAGAGTTGTTTTCAGAAATGCGCGAGGTT 1053
1357 hAla...LeuSerMetSerAsn...ProHisIle 1366
1052 GTR...CGCGTATGAGTATTTGTCGCGC... 1026
1367 SerHisSerHisValProAlaIleValSerGlyAlaPheSerSer 1383
1025 ...TCGGCGCACCCAGCCGACAGCTTTG...CTGCGC 989
1383 rAlaAlaIleGlyThSerThProAsnSerGlyLeuSerThLeuAlav 1400
988 CTTCCTGATACGGAATCTGATGTTGATGCTCCCAATAATCTGCG 939
1400 alPhrSerLeuSerThr...SerIle 1407
938 GCGCGCTGTAATCCGCGCTTCATACGAGCAGGAATACGCGGTT 889
1408 AlaPro...GlnProHisSerHisPheProGlnSerThGlnMe 1421
888 G...TCTGTG... 882
1421 tLeuProGlnSerGlyAsnPheSerSerValSerHisLeuThrPthN 1438
881 ...TCACCAATTCGCGCGCAGTATTTGCGACTTTCGCA 843
1438 tSproMetSerSerGlnAsnGlnProMetValArgGlySer... 1452
842 CCCAAACGCTGACCAAGACGCGCGCTTTCAGACTTTCGAGAACGCTG 793
1453 ...ThLeuTySerGlnSerSerAlaAlaIleThAlaProProse 1467
792 GGCATACAGCGCGCTGTTTCAGAGCGCTTTCGCAACGAGCGCA 743
1467 rAlaAlaAla...AlaValSerAsnPheThProse 1478
742 TGGTAATTCATCTGTGTAATGATGTCACACAGGTTTATTCGCGCG 693
1478 ervaLeu... 1480
692 ACCGCTGATGAAATGATGCTGTCACATCAACCGCAGAGATCGG 643
1480 ... 1480
642 GCGCGCAATTCATGATGTTTCGATGTCGACGATTTTCAGAGCGCAGT 593
1481 ...AlaValGlnSerLeuThrThAlaValThSerSerSer 1494
592 CTGCGCCAGCTGCTTACAAACATGATTTGCGTGGTC... 552

```

1494   .....AAAGCGCTCAATACCAACAGCCGGCGTTTGAATCCTCGGC    511
      |||.....:|||||||:|||||||:|||||||:|||||||:|||||||:
      SerProSerThrLeuSerSerSerValIleGlnIlyValIleSerPro    1510
1511   LysGlyGluSerProCysAnlysAspArgSpSerSerTyrlSerSerPr    1527
      :||:..|||:..|||:..|||:..|||:..|||:..|||:..|||:..
      ..:::~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::
      oAlasAlaIalValIalThrThrCysAlaProThrThrProIleValSer    1543
1544   TGCATTGACGAAGAATGGCAAAGCGCTGGCATGACGACGAGAATTGTG    411
      .....SergIlySerAlaIalArgProThrProProIleu    1553
1554   CTGAAAC.....GGAGSGTGCGCACGCCACTCCACA    379
      |||.....|||:|||||||:|||||||:|||||||:|||||||:
      SerAsnGlySerMetLylIleYlValAlaAla.SerThra    1570
1570   ACCGATTGGATCAGTGGTCGGCGGCACCTTTCGCCCTTAAGTTGGCA    329
      |||.....|||:|||||||:|||||||:|||||||:|||||||:
      lAArg.....SerSerCysAsnAlaIle..SerProLeuSerIleProA    1584
1584   GCGCTTCAGT.....GCGTAGCGCTTCAAC    303
      :|||||||:..:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::
      lathrAlaIlylIleHlsValSerAlaThraSnProSerPheGlnSerSer    1600
1601   TCGATTTCGTGTCCT.....                285
      |||.....
      SerTyrlPheProThrProLeuAlaProProSerSerProSerProAl    1617
1617   ATHTSerSerAlaAlaIleIleSerSerSerAlaSerGlnPheAsnProA    1634
      |||:|||||||:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::
      CGGTATCGCGCGCATTTTGGCTGCAACGGCGGCACATAAAC    242
      :~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::
      lavalSerHisSerMetSerSerIleValIalThrAlaIyAlaThrThr    1630
200   ACCAGCGCCGGATTCTTTTGTCTTCAACGACGACTGGCCTTTTGTGAC    151
      |||:|||||:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::
      ThrThr.....AlaSerSerValThr.....        1657
1657   GGCATCGCCTTCCTTGACTTCATCGAGGGCGCATACCGGCAATVTCTT    101
      ::|||:|||||:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::
      GlmProSerValAlaAlaIleSerAsnProValIThraSnThrProH    1658
100   CCCCAAGCAACGCGACTTCGGA    78
      |||:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::
      lAsProPheSerAlaGluSerLeu    1673

seq_name= sp_plant:Q9SUVO

seq_documentation_block:
ID       Q9SUVO          PRELIMINARY;           PRT;      857 AA.
AC       Q9SUVO;
DT       01-MAY-2000 (TREMBLrel. 13, Created)
DT       01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE       01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE       HYPOHERETICAL_96.9 KDA PROTEIN.
GN       FGB4_120 OR AF4G32420.
OS       Arabidopsis thaliana (Mouse-ear cress).
OC       Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC       Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC       eurosilvites II; Brassicales; Brassicaceae; Arabidopsids.
ON       NCBI_Taxid-3702;
RX       [1]
RP       SEQUENCE FROM N.A.
RA       Bevan M., Terry N., Ardiles W., Buysshaert C., Dasseyille R.,
RA       De Clerck R., De Keyser A., Neyt P., Roupe P., Van Den Daele H.,
RA       Valleron J., Gielens K., Van Montagu M., Hohneisel J., Meves H.W.,
RA       Mayer R.F.X., Lemcke K., Schneller B.C.
RL       Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
```

RN [2] EU Arabidopsis FROM N.A.
 RP EU Arabidopsis sequencing project;
 RA Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Terryn N., Ardles W., Buysshert C., Dasseville R., De Clerck R.,
 RA De Keyser A., Neyt P., Rouze P., Van Den Daele H., Villarroel R.,
 RA Gielen J., Van Montagu M., Mewes H.W., Lemcke K., Mayer K.F.X.,
 RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL034567; CAA25656.1; -
 DR EMBL; AL161581; CAB79959.1; -
 DR HSSP; P05092; 2CPL,
 DR InterPro; IPR002130: CSA_PPIase.
 DR Pfam; PF00160: pro_lisomerase; 2.
 DR PRINTS; PR00135; CSAPIISMRASE.
 DR PROSITE; PS50072; CSA_PPIASE_2; 1.
 RW Hypothetical protein.
 SO SEQUENCE 857 AA; 96854 MW; 17EAFa842D58D972 CRC64;

```
alignment_scores:
    Quality: 118.00
    Ratio: 0.663
    Percent Similarity: 45.293
    Length: 393
    Gaps: 22
    Percent Identity: 25.700
```

```

Align seg 1/1 to: Q9SUW0 from: 1 to: 857

24  AAACCTGCCCATCGCGGGGCACACCGGACGAACCGCTTACGACGGCCCGG 73
   :::: ||| ||||||| :::: ||| :::: ||||||| :::
520 ArgSerAlaSerArgGlySerLeuGlyArgGlyProLeuArgArgSer 536
74  CCATTACCGAAGTGCCTGTGCTTGGCCAGCAAAATATGCGGTATGCGGCC 123
   : ..... ||||| ::::|
536 : ..... ArgArgSerProSerArgSerPro 545
124 TCGATGGAAGTCACGAGAAGCGATGCCGTCAAAAAAGCCAAATGCTGTT 173
   ::::| ||| ||||| ::::| ||| :::
545 aArgSerSer..... ArgArgSerLeuSerArgSerProLeuGlnLeu 559
174 TGAAGACAAAAAGAAATCCGGGCGGTGGTTTACTGCGCCGCGCTTCAGGCA 223
   ||| ||||| ::::| ||| |||
560 ..... SerArgArgSerLeuSerArgSerProThrArgLe 571
224 AAATCGCCGCGAT..... TCACCGTGGCGAANAAGCGCGTACTTTCAG 264
   ::::| ||||| ::::| ||||| ||||| ::::| ||| |||
571 userArgArgSerLeuSerArgSerProLeuArgSerProArgLySerV 588
265 TCAATCTGTGATTCGCGTTGAAGCACAACGAGAAATGAGATTCAACGCTA 314
   ||||| ::::| ||| ::::| |||||
588 AlserArgSer..ProValArgSerSerArgLySer..... 599
315 CGCACTGAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGCCGCAACC 364
599 ..... 599
365 TGATTCATTCGGTTGTGGACHTGGCGCTCGCACCGCTCCGTTACGA.. 412
600 ..... ::::| ||||| ||||| :::
ValSerArgSerProValArgSerSerArg 609
413 .AAATTCTGCGCTGCATCCGACGCGCTTCGCATCTTCGTCAATGCAT 461
   ::::| ||| ||||| ::::| :::
609 gArgArgGlySerArgSerProValArgSerSerArgLySerValSerA 626
462 GGACACCAATCGCGTGGCGTCCGACCCCTAAGGTCAATTATCAAGAAGCGC 511
   ::::| ||||| ::::| ||||| ::::| |||

```

```

626 rgSerProIleArgLeu.....SerArgArgSerIleSerArgSerPro 640
512 CGAGGATTTCAACGGCGCTGTGGTATGAGCG.....548
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
641 lIeArgLeuSerArgArgSerIleSerArgSerProValArgIleArg 657
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
549 .....TTTGACCGAACGCAAAATCCATGTTGTAAAGCAGCTGGCCGCA 593
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
657 rglIeArgLeuSerArgSerProValProAla.....ArgArg 668
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
594 CGTGGCTGTGAATGCTGCCACATGCAACATGATTCGGCGGCC 643
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
669 ArgSerValArgPro.....ArgSerPr 676
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
644 CGCATCCCTCCGGTTGAGTGGACGACCATTCATTCAGCGCGCTC 693
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
676 oProProAspArgArgArgSerIleSerArgSerProAlaSerProAla 693
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
694 GCGCGCAATTAACCGCTGGACCATTCATTCAGATGTAATTCAGAT 743
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
693 rglIeArgLeuArgIleArgIlePheSerGln.....ArgPheSerTy 706
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
744 TGCCCGTTGTTGCAACAGCGCGCTGTGAACACGAGCGCGTGAATGGCC 793
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
707 .....AlaArgArgTyArgThrSerPr 714
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
794 TAGGTGTTCTCAAGTCACAAACCGCGCTGCTGCTACCGCTTTGGGT 843
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
714 O.....SerProAspArgSerPro.....TyArgPheSer 724
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
844 GCGAATGATCGCAATTAAGTGGCGGCGAATGTTGACACACAGACACCG 893
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
725 .....AspArgSerAsp 728
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
894 CGTATTCGCGTTCGGTATTAAGCGCGCGATTAACACAGCGCGACG 943
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
729 ArgAsp.....ArgPheArgSerArgArgArgPheSerProSerArgPheArg 744
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
944 A.....TTATTTGGGACGCTACACATTCATTCCTGTTATCGAAGA 987
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
744 gSerProLeuArgIleArgThrProSerSerMetLeuArgIleArgArg 761
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
988 GCGCGCAGCAAGAGCTGTGGCTGGCTGGCGCCGACCGGACAAATA 1037
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
761 rg.....SerArgSerValSerProIleUcystTyArgAsnArg... 774
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1038 CTCATCAGCGGTACACCT.....CGGCATTTCTCTGA 1072
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
775 .....ArgTySerArgSerProIleArgSerArgSerProTyArgIle 790
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1073 AAACCAACTCTTCAGTTCAACACAGCGCTCAACGCGCGACCGCGCC 1122
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
790 sArgArgSerProSerAlaSerIleSerLeuSerProSerArgSerArg 807
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1123 ATGTGGCGATGTGTTACTACGACG 1148
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
807 erArgSerTySerTySerTySer 815
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: sp_bacteria:093LB1
seq_documentation block:
ID 093LB1 PRELIMINARY; PRT: 299 AA.
AC 093LB1:
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PUTATIVE DNA-BINDING PROTEIN.
OS Bifidobacterium animalis (Bifidobacterium lactis).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=28025;
RN 11
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-ATCC 27536:
RA Gonzalez Yara A., Rossi M., Altomare L., Gottl R., Matteuzzi D.:
RT "Effect of pH and culture redox potential on the continuous culture of
RT Bifidobacterium animalis ATCC 27536, studies on stability of
RT recombinant plasmids and characterization of the insertion sequence
RT IS1999."
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ18089; CAC41624.1; -.
KW DNA-binding.
SQ SEQUENCE 299 AA; 34340 MW; B71D96C9ECON7A03 CRC64;

alignment_scores:
  Quality: 115.50      Length: 385
  Ratio: 0.726        Gaps: 21
  Percent Similarity: 41.299      Percent Identity: 23.896

alignment_block:
US-09-303-518d-125 x 093LB1 ..

Align seg 1/1 to: 093LB1 from: 1 to: 299

81 CGAAGTCGCGTTCGTTGGGAGATATGCGCGATGCGCCCTCGATGA 130
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
32 ArgAlaArgValAlaProArgArgValArgArg.....ProVal..... 44
131 AAGTCAGAGAGCGATGCGCTCAAAAAAGCGCAAGT...GCTGTTGAA 177
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
45 .....ArgAspArgAspArgArgProAspProAlaLeuArg 57
178 GACAA.....AAGAAATCCGGCGCTGGTTCATGCGCGCGCTCAG 221
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
57 rglIeArgLeuArgIleArgIleArgIleArgIleArgIleArgIle 67
222 CAATAATCGCGCGATTCACCGTGGCGAAGAGCGCGCTACTGAGTCG 271
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
68 .....GlyProPheArgProGlyAlaSerArgMetAspAl 79
272 TGATTCGCGTTGAGGCAACGACGAAATCGATTTGAACGCTAGCACCT 321
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
79 aglAlaArgValArgValArgValHis..... 86
322 GAAGCGCTGGCAAACTTAAGCGCGAGAAAGTGGCGCGCAACCTGATCA 371
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
87 .....LysThrArgArgGlyGly..... 92
372 ATCCGTTTGTGAGTGGCTGGCGACCGCTCCGTTACAGAAATTCCTG 421
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
93 .....ArgHisGlyArgVal 97
422 CGTGCATGCGGAGCGGTTGCGCATCTTGCATTCGATGACACCAAT 471
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
97 lArgArgLeuGlnAlaGlyArgArgIleArgArgAlaPro..... 110
472 CGGCTGGCTGCGACCGCTACGTCATTAACAAAGAACCGCGGAGATTT 521
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
111 .....ArgGlyArgGlyHisArgProValProAlaArgValAlaThr 123
522 CAACGCGCGCTGTGTGATTAAGCCGTTGACGCAACCAAAATCCATG 571
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
124 GlyArgArgPro.....ArgHisGlnGlyAlaLe 133
572 TTGTAGGAGCAGTGGCGGACAGTGGCGCTGTGMAAATGCTGCACATC 621
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
133 uProProAlaThrArgArgHisArgGlnAla.....ArgCysGlnGly 148
622 GAACACATGAATTCG.....CGGCGCGCATTCGCGCTGTTGAG 662
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
148 rglIeArgLeuGlnAlaArgProAlaAspGlnArg..... 161
663 TGGCAGCGACATTCATTCATCGAGCGGCTGGCGGAATAAACCGTGT 712
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
162 .....ArgIle 163

```



```

1075 TTTTGGAGAAATGGCCGAGGTTGACGC..... 1047
      |||||:
807 .....ValVallysglyThrThrThr 816
1046 .....GTGATGAGATTTTGTCCGCGTCCGCGC 1018
      |||:
817 LeuGlyThrValThrGlyThrValSerThrSerLeuAlaGlyAla 833
1017 AACCCAGCCGAACAGCTTTGTGGCGGCTTCTGCATAC..GGAATC 969
      ||:
833 aHisSerThrSerAlaSerLeuAlaThrProIleThrLeuGlyThrI 850
968 TGATGTGGTAGCGTCCCAATAATC..... 943
      ||:
850 leaIaThrLeuSerSerGlnValIleAsnProThrAlaIleThrValSer 866
943 ..... 943
867 AlaAlaGlnThrThrLeuThrAlaAlaGlyLeuThrThrProThrI 883
943 ..... 943
883 eThrMetGlnProValSerGlnProThrGlnValThrLeuIleThrAla 900
942 .....GTGGCGGCTTG.. 931
900 roserGlyValAlaAlaGlnProValIleAsnLeuProValSerIleu 916
930 TGTAAATCGCGCGTCAATACGAC..... 904
      ||:
917 AlaSerProThrThrGlnGlnProThrAlaThrValThrIleAlaSp 933
903 .....GGAATCAGCGGTTGTCTGTCAAC 876
      ||:
933 rGlyGlnGlyAspValGlnProGlyThrValThrLeuValCysSerAsn 950
875 AATTCGCCGAGATTTTGGCATCTTGGCACCACCAACGCGTACGCA 826
      ||:
950 roProCysGlnThrHis.....GlnThrGlyThrThr 960
825 GAGCGCGGTTGTGACTTGAACAACACCTAGGCAATCAC.....GC 782
      ||:
961 AsnThrAlaThrThrThrValAlaValAlaLeuGlyGlnSProGlnPr 977
781 GCTCGGTTTCAGAGCGCTTGTCAACAA..ACGGCC..... 745
      ||:
977 oThrGlnValGlnPheValCysAspArgGlnGlnThrAlaAlaSerLeu 994
744 .....AATGTAATTAATCATCTGATTAATGAT 718
      ||:
994 alThrSerAlaValGlyGlnGlnAsnGlyAsnValAlaValGlyCysSer 1010
717 GGTCCACACGGTTTATTCGCGCGCAGCGCTCATGAATGATGCG 668
      ||:
1011 AsnProPro.....CysGlnThrHisGlyThrGlyThrThrAla 1025
667 TGCCATCAACACCGCAGATGCGCGCGCGGATTCATGT..... 625
      ||:
1025 aThrThrAlaThrSerAsnMet..AlaGlyGlnHisGlyCysSerAsn 1041
624 .....TTGCATGTTGCAACATTTTCAGACGCGCATCTGC 589
      ||:
1041 roProCysGlnThrHisGlyThrGlyThrThr..SerThrAlaThrThrAl 1057
588 GCGAGTCCCTTACAAACATGATTTGGTGGGTGAACGCGCTCAATA 539
      ||:
1057 aMetSerSerMetGlyThrGlyGlnGlnAlaArgAspThrAlaArg...Th 1073
538 CCAACAGCGCG.....CGTTGAATCTCGCGCGCTTCTTGATA 498
      ||:
1073 hrAsnThrProThrValAlaArgIleThrValAlaProGlyAlaLeuGly 1089
497 ATGACCGTAGGTCGCGCA.....GCCAGCGGATTT 469

```

```

1090 ArgValAlaGlyThrValValysProGlnCysGlnThrGlnGlnThrAsn 1106
      ||:
468 GGTGCCATTCGATTTGACGAAGATGGCAACGCGCTCGCATCGCGCAG 419
      ||:
1106 tThrThrThrThrMetThrValGlnAlaThrGlnAlaProCysSerAlaG 1123
418 GAATTTGCTGACGACGAGCGGTG.....CGCAGCGCATTCACAAACCG 375
      ||:
1123 LyrProLeuLeuAlaGlyProSerValAlaLeuGlnSerGlySerHisSer 1139
374 GATTGGATCAGGTTGCGCGCGCAGCTTCTCGCAGCGGTAAGTTTCCGCGC 325
      ||:
1140 AlaPheValGlnLeuAlaLeuProSerValAlaArgValGlyLeuSerGly 1156
324 TTCAGTGGCGTACGCTTCAACATCGATTTGCTGCTTGCCTTCACGCGCA 275
      ||:
1156 oSerSer.....LysAspMetProThrGlyArgGlnProGlyThrThr 1171
274 TCACGACTGACTGAAGTACGCGCGCTTTCGCAACGCGTGAATCGCGCGCAT 225
      ||:
1171 IsThrThrThrThrAsnThrProThrThrThrArgSerIle...MetVal 1186
224 TTGCTGTAACCGCGCGCAGTAAACACACGCGCGGATTTCTTTGTCTTC 175
      ||:
1187 AlaGlyGlyLeuGlyAlaAlaAlaArgValAlaPro..... 1197
174 AACACACACTTGCGCTTTTGGACGCA...TCGCGTTCCTTGACTTCA 128
      ||:
1198 .ThrSerThrThrGlySerLeuGlnAlaSerSerProSerSerThrMet 1214
127 TCGAGGGGCGATACCGCGCATTTCTTCGCAAC..... 93
1214 hMetThrAlaLeuGlnAlaLeuLeuCysProSerAlaThrValThrGln 1230
92 .....AA 91
1231 ValCysSerAsnProProCysGlnThrHisGlyThrGlyThrThrAsn 1247
90 CCGCAGTTCGGTAAATGCCGCGCGCTGTAACGCGCTTGTCCGCTCGC 41
      ||:
1247 rAlaThrThrSerAsnAlaGlySerAlaGlnArgValCysSerAsnPro 1264
40 CC 39
1264 to 1264
seq_name: sp_mammal:029071
seq_documentation_block:
ID 029071 PRELIMINARY; PRT; 528 AA.
AC 029071:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE GASTRIC MUCIN (FRAGMENT).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9825;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE=GASTRIC MUCOSA;
RX MEDLINE=95275264; PubMed=775593;
RA Turner B.S., Bhaskar K.R., Hadjopoulos-cladaras M., Specian R.D.,
RA Lamont J.T.,
RT "Isolation and characterization of cDNA clones encoding pig gastric
RT mucin."
RL Biochem. J. 308:89-96(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE=GASTRIC MUCOSA;
RX MEDLINE=94102478; PubMed=7506218;

```

708	GGTTTTATTTCGCGCCGACCGGCTCGATCAATAATGATGCGCTGGCACCA	659
	: : : :	
328	rserGlySerAlaProThrThrSer.....AlaThrSerValG	341
	: : : : : :	
658	AACCGGCAGATGCGGCGCGCAATTCATGTTTCGATGTTGGACGA	609
341	lInProSerSerSerSerProProlIleSerSerThrIleSerValGln	357
	: : : : : :	
608	TTTTAGAGGACGCTTCGCCAGCTGCTTACAAACATGATTTTCG	559
	: : : : : :	
358	ProSerSerSerSerSerProThrThrSerThrThrSerVal.....	372
558	TTGGGTCAACAGCGCTCAATACCAACAGCGCGGCTTGAATCTTCGGCG	510
	: : : : : :	
373GlnProSerSerSerGlySerAlaP	381
	: : : : : :	
509GCTTCTTGAATATGACCGTAGGCTGGCGACCGCGGATTTGGTG	465
	: : : : : :	
381	rOthrThrSerAlaThrSerValGlnProSerSerSerSerSerValPro	397
464	TTCATCGCATTTGACGAAGATGGGACAGCGCTCGGCATCGACGCGAGAAT	415
	: : : : : : : : :	
398	ThrThrSerAlaThrSerValArgSerSerSerSerSerThrProI	414
	: : : : : : : : :	
414	TTTGCTGAACGAGCGGTCGGCAGCGCACTCACAAACCGGATTTGATCA	365
	: : : : : : : : :	
414	eProThrThrThrSerValGlnProSer.....	423
	: : : : : : : : :	
364	GGTTCGCGCCACTTCTTCGCCGCTTAAGTTTCCAGCGCTTCAGGTGCG	315
	: : : : : : : : :	
424SerSerSerSerValProThrThrSerAlaThrSerVal	436
	: : : : : : : : :	
314	TAGCGTTCAACACTGCATTTCTGTCG...TTGCTTCAACGCGCAATCACGAC	268
	: : : : : : : : :	
437	GlnThrSerSerSerSerThrProIleProSerThrThrSerValGln	453
	: : : : : : : : :	
267	TGACTGAAGTACGCGCTTTTCGCCACGCGTGAATTCGGCGATTTGGCTG	218
	: : : : : : : : :	
453	nProSerSerSerSerSerAlaPro.....	461
	: : : : : : : : :	
217	AAGCGGCGCGATTAACACCAACGCGCGGATTTTGTCTTCAACAGC	168
	: : : : : : : : :	
462ThrThr.....SerAlaThrSer	467
	: : : : : : : : :	
167	ACTTGCCCTTTTTCAGCGCATCGCTTCTTACCTTTCATCATCAGGCGCG	118
	: : : : : : : : :	
468	ValGlnProSerSerSerSerSerProIleSerSerThrIleSerVal	484
	: : : : : : : : :	
117	CATACGGCATATCTTTCGCCAACCAACGCGACTTCGGTATG.....	75
	: : : : : : : : :	
484	IglnProSerSerSerSerSerProThrThrSerThrThrSerValG	501
	: : : : : : : : :	
74GCCGGCGCGTCGTAAACGCGCTTGCTCGGCTG	42
	: : : : : : : : :	
501	lInProSerSerSerSerGlySerAlaProThrThrSerAlaThrSerValGln	517
	: : : : : : : : :	
41	CCCGCGATGGCAGGTTTACAGACT	18
	: : : : : : : : :	
518	ProSerSerSerSerSerProPro	525
	: : : : : : : : :	
seq_name:	sp_Bacteria:Q93SJB8	
seq_documentation_block:		
ID	Q93SJB8	PRELIMINARY; PRT; 575 AA.
AC	Q93SJB8;	
DT	01-DEC-2001 (TReMBLrel, 19, Created)	
DT	01-DEC-2001 (TReMBLrel, 19, Last sequence update)	
DT	01-DEC-2001 (TReMBLrel, 19, Last annotation update)	
DE	USC3-5P.	
OS	Myxococcus xanthus.	
OC	Bacteria; Proteobacteria; delta subdivision; Myxobacteria	
OC	Myxococcales; Cystobacteriineae; Myxococcaceae; Myxococcus	
OX	NCBI_TaxID=34;	

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CECT 422;
 RA Poza M., Sileiro C., Villa T.G.;
 RT "Clone USC3 from *Mycococcus xanthus* CECT 422 strain."
 RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases
 DR EMBL; AY033403; AKK49004.1; -
 SQ SEQUENCE 575 AA; 62586 MW; D2C19B7G363C5634 CRC64;

alignment_scores:

Quality:	112.00	Length:	573
Ratio:	0.521	Gaps:	29
Percent Similarity:	37.522	Percent Identity:	21.815

alignment_block:

US-09-303-518D-125 x Q93SJ8

Align seg 1/1 to: Q93SU8 from: 1 to: 575

9 AATCAAAAAGACGTAAACCTGCCCATTCGGCGGCAGACCGGACGCAAC.. 56
 31 ***Ala***Argmet***ProAlaAsnAlaHisIuIneSerAlaAla** 47
 57CGTTACGACGGCCCGGCATATTACCGAAGTCGCTTG 93
 47 *GlyAlaArg**ArgProGlnArgProAla**ArgSerAlaAlaIleA 64
 94 CTTGG.....CGAAGAATATGCCGGATAGCCGCCCTC...GATGAA 131
 64 LaSerThr***ProArgSer***AlaLysAlaGlyThrLeuAlaThrPro 80
 132 AGTCAGGAAGGCGGATGCGTCAAAAAGGCGCAAGTCTGTGGAGACA 161
 81 ThrHisIysSerSerAsnArgSerLysArgProAsnPro....ArgI 95
 182 AAAAGAATCCGGCGCGTGTCTTACTGCGCGCGCTTCAGCAAAATGCGC 231
 95 ngInaArgArgIAlaArgAsnAla.....SerArg 105
 232 GCATTCACCGTGGCGCAAAAGCCCGTACTTCAGTCACTGATGATCCGT 281
 105 Iaasp.....GAAAGG.....GTT..... 111
 282 TGAAGGCACGACGAATGAGTGTGAAGCTTCGCGACCGTGAAGCCGTG 331
 115AlaAlaGlyThrAlaArgValProProAlaG 125
 332 CAACCTTAAGCGCGCA..... 347
 125 yProAlaArgAlaArgArgSerProSerArgSerArgAsnArgTrpArgLeuA 142
 348AGAATGGCGCGCAACCGCA.....TCCATCCGGTT 379
 142 rGcysGlyThrArgProAlaProAlaProAlaProTyrPalaAlaMetThrThrVal 158
 380 TGTGACGTCGCGTGGCACCCGCTCGCTTCAGCAAAA...TTCCGTCGGTC 426
 159 AlaGlySerAlaProAlaProAlaProAlaArgProAlaArgAlaValLeuProAr 175
 427 GATGGCG.....AGCCGTTCCGCATCTTCGTCAATGC 458
 175 galProLeuGlnArgThrAlaAlaProAlaArgAlaGlyProSerGly_Cys 191
 459 GATGACACCAATCCGCTGGCTGCGCGACCGCAACGGT..... 494
 192 ProAlaAlaArgCysProGlySerCysProProGlyAlaGlyArgSerSe 208
 495CATTCATAAGAGCCGCCGAGGATTCATAACGCGGCGCTG 534
 208 rTrpPrometGlnTyrAlaArgAlaArgArgValArgArgProSerPro. 224
 535 TTGGATATGAGCCGTTTACCGCAAGCAAAATCCATGTGTTGTGAAGCAGC 584

[illegible]


```

297 lYThSerThrThr.....ThraSerserCluster 307
158 TTTTGAGCGATCGCTTCCTTCAATCAGAGGCGGCAATACCGGC 109
   |||:|||||:
308 AspleuA1atThrThr.....Prova 314
108 ATATCTTCG...CCAGCAACGCGACTGCGTAATGCGCGG..... 69
   |||:|||||: |||:|||||: |||:|||||:
314 lYtSerSerMetProPheserThrThrlyValThSerGlySerAlaI 331
68 .....CCGTCGTAAACGGCTTCCTCCGCTCCGCC 39
   |||:|||||: |||:|||||: |||:|||||:
331 lelleProaSPHisnGlySerSerValleuPro 342

seq_name: sp_invertebrate:Q9U0V2

seq_documentation_block:
ID Q9U0V2 PRELIMINARY; PRT: 2207 AA.
AC Q9U0V2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE POSSIBLE MUS308 HOMOLOG (FRAGMENT).
GN L8342.01.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_Taxid=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Masny D., Purrelle B., Goffeau A., Ivens A.C., Lawson D., Quail M.,
RA Rajendream M.A., Bartell B.G.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.;
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
DR EMBL; AL122012; CAB58415.1; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00270; DEAD.1.
DR Pfam; PF00271; Helicase_C.1.
DR SMART; SM00490; Helicase.1.
KW ATP-binding; Helicase.
FT NON_TER 2207 2207
SQ SEQUENCE 2207 AA; 229262 MW; F461565DA0017B0C CRC64;

alignment_scores:
Quality: 112.00 Length: 457
Ratio: 0.599 Gaps: 24
Percent Similarity: 40.919 Percent Identity: 22.976

alignment_block:
US-09-303-518D-125/rev x Q9U0V2

Align seg 1/1 to: Q9U0V2 from: 1 to: 2207

1232 CAACCAATGCGTCGCGTGTCCGATCCGACGATTAATCGGCGAA 1183
   |||:|||||: |||:|||||: |||:|||||:
813 GluProaTrgLeuGlySerValAlaThrThr..... 824
1182 AAGAGGGTGGCGATATCCAGGCGATCAGCGC.....T 1145
   |||:|||||: |||:|||||: |||:|||||:
825 .....ThelGlnAlaSerAlaGlySerAlaArgAsnValAlaAla 839
1144 CGTAAGTACCAATCGCACCATGCGCGCGCGCGCTTGAGCGGTG 1095
   |||:|||||: |||:|||||: |||:|||||:
839 erValAlaGluProAlaThrProValSerSerProSerGluAla... 854

1094 TTGAAGTTGAGAGATTGTTTTCAGCAAAATGGCCAGGCTGTACGCGT 1045
854 ..... 854
1044 GATGAGATATTTGTCGCGTGGCGGCGACCGGACAGCTCTTTCG 995
   |||:|||||: |||:|||||: |||:|||||:
855 .....CysAlaThrThrHisPro...ProaSPleu 864
994 TGGCGCTTCTTCGATACGGAATGATGATGTAGCGTCCCAATAA 945
   |||:|||||: |||:|||||: |||:|||||:
864 rPluProSerLeuProThrGluHis..... 872
944 TCGTGGCGGCTGTGTAATCGCGGCTTCATACCGAACCGGAATCAC 895
   |||:|||||: |||:|||||: |||:|||||:
873 ValSerProProCys.....AspProLeuLeu 882
894 GCGGTGTCTGTGTCACCAATTCGCCGAGTAATTCGATCTTCG 845
   |||:|||||: |||:|||||: |||:|||||:
882 lArgValSer...SerGlnGlyGlyAlaGluGlyCys..... 894
844 CACCAAAACGTAACGAGAGCGCGGCTTGTGACTTGAGAACACACT 795
894 ..... 894
794 AGGCAATCACGCGCTGCTGTTCAGACGCGCTTGCAAAACAGCGC 745
   |||:|||||: |||:|||||: |||:|||||:
895 .....AlaAlaGlnSerValPheArgMetGlyValAlaAlaProThrPr 909
744 ATGTAATTAATCATCTTGATTAATGATGTGTCACAGGTTTATTCGCGC 695
   |||:|||||: |||:|||||: |||:|||||:
909 oThrSerLeuGlySer.....AspValLeuSerSera 920
694 CGACGCGCTCGATGAATGAATGCTGTCACCTCAACCGCAGAGATGC 645
   |||:|||||: |||:|||||: |||:|||||:
920 lathrGly.....ValSerAlaLeuSerAlaAlaSPnla 931
644 GGGCGCCCAATTCATGCTGTTTCGATGTGCA..... 612
932 AlaProProLeuSerAspLeuHisValThrAlaLeuProTyLeuAlaThr 948
611 .GCATTTTCAGAGCGAGCTGCGCGCGCTTTCG.....GCT 578
   |||:|||||: |||:|||||: |||:|||||:
948 rAlaAlaAlaGlyGlySerGlyAlaProAlaValProAlaArgProAla 965
577 TCAACAACATGATTTGCTGCTGCTCAACGCTCAATACCAACAGCGCG 528
   |||:|||||: |||:|||||: |||:|||||:
965 rGThrCysPheThrLeuHisSerAlaAlaArgThrThrGlyThr..... 979
527 GCTTGAATCCTCGCGGCTTTCG.....ATAATGAC 493
   |||:|||||: |||:|||||: |||:|||||:
980 .....LeuSerSerSerAlaAlaAlaThrSerThrGlnGluProThrThrAl 995
492 CGTAGGTCGCGACCGACGAGATTTGCTTCATCGCATTCAGCAAGATGG 443
   |||:|||||: |||:|||||: |||:|||||:
995 aAlaAlaSerAlaSerThrGlyAlaValProProAlaLeuThrGluAla 1012
442 CG..... 441
1012 erTyTrLeuSerAlaGluLysCysPheArgGlyGlyThrAlaGluLeu 1028
440 .....AAGGCTCGGATCGACGCGCA..... 417
1029 GlnGlnLeuProGlnHisGlyAlaAlaAlaSerProGlnGlyThrGluGlyLe 1045
416 .....ATTTCGTGACGAGCGGTGCGACGCGCAGTC..... 384
1045 uMetValSerLeuAspGly...ValAlaSerAlaThrProSerAspThrAla 1061
383 .CACAAACGGAATGATGATGCTGCGCGCATCTTTCGCGCTTAA 336
   |||:|||||: |||:|||||: |||:|||||:
1061 rGlySerMetProSerSerMetGluSerArgHisThrThrGlyCysAlaGln 1077

```


[illegible]

OC Actinomycetales; Micrococciaceae; Microbacteriaceae; Microbacterium.
 OK NCBI_TaxID=42460;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15354;
 RX MEDLINE=96257250; PubMed=8654988;
 RA Striebel H.M., Seebler S., Jarsch M., Kessler C.;
 RT "Cloning and characterization of the Mami restriction-modification
 RL system from Microbacterium ammoniaphilum in Escherichia coli.";
 DR EMBL: X79027; CA55649.1; -
 KW Hypothetical protein.
 FT NON_TER
 SO SEQUENCE 529 AA; 58716 MW; 5DC1216237DF772D CRC64;

alignment_scores:
 Quality: 111.00 Length: 526
 Ratio: 0.491 Gaps: 29
 Percent Similarity: 42.966 Percent Identity: 24.715

alignment_block:

US-09-303-518D-125 x P94909 ..

Align seg 1/1 to: P94909 from: 1 to: 529

```

21 TCCTAACCTGGCCATCGCGGCGAGACCGGAGCGGCTTNGA...CG 67
   ::::::::::::::::::::
49 AAlaySProgluArgAlaIuProArgGlyThrlleIuProArg 65
   ::::::::::::::::::::
68 GCCCGGCAATTCGGAAGTGGCTGCGGAGAGAAATATGCGGTAG 117
   |::::::::::::::::::
65 gGlySerHisTyrAlaIuArgIuProAlaIuArg..... 77
118 CGCGCCATCGAAGTCAAGAGGCGATGCCGCAAAAAGGCGCAAGT 167
   ::::::::::::::::::::
78 .....ArgIuAlaHisProIeu 83
168 GCTGTTGAAGACAAAGAAATCGCGGCGTGTCTTACTCGCGCGCTT 217
   ::::::::::::::::::::
84 GlnATrPronIsgIuIuProgluAlaArgIuGlu.....LeuProva 98
218 CAGCGAAATCGCGGCGATTCACCGGCGGAGAAACGCGTCTAGTCA 267
   ::::::::::::::::::::
98 lArgAspArgAlaIuArgIuGlyArgIuGluIuProAlaArgProAla 115
268 CTGCGATGTCGGT..... 281
115 lArgAspAlaArgIuHisIeuArgIuAlaHisProAlaArg 131
282 .....TGAAGCAAGCAAGAAATCGAGTTGAAACGCGTACGCGCTG 322
132 GAluAlaIuArgIuHisArgIuHisArg..... 141
323 AAGCGCTGGCAAACTTAAGCG.....CGAAGAAGTGGCGGCAACGCTG 366
   ::::::::::::::::::::
142 AAlaArgIuGluArgIuArgIuProArgGlnIuAlaIuProArgIuHis 158
367 ATCCAAATCGGTTTGTGACTGCGCTGGCGACCGCGTCCGTTCAAGCAAAAT 416
   ||::::::::::::::::::
158 lSProArgIuArgIuArgAspArgAlaIuArgIuProgluIuHisArg... 173
417 TCCTGCCGCTGCATGCGGAGCC.....GTTGC 442
174 .....ArgATrArgAlaIuArgIuArgIuAlaIuArgProgluAlaIu 188
443 CCAGCTTCGCAATGCGATGAGACACCAATCGCGTGGCTGGCGAGCC.... 488
188 gHis.....GlyAspArgGlnHisArgAlaAspProArgATrProArg 203
489 .....TAGGTCATTAAT.....CAAGA 506
203 sPProAlaIuArgIuHisHisProIeuArgIuGlnIuGlnIuGlnIu 219
```

```

507 AGCCGC.....CGAGATTTCAACCGCGCTGTGGTATTGA 544
   ::::::::::::::::::::
220 AAlaArgProIeuArgAlaArgGlyGlnIuGlnIuProgluGly..... 234
545 GCCGTTTTCAGCAGCAAGCAAAATCCATGTTTGAAGGACGCTGGCGC.... 590
   ||::::::::::::::::::
235 .ProAlaIuArg.....HisProAlaIuGlnIuAspGluAspArgPro 249
591 .....AGAGTGGCGCTGTGAATGC 611
249 rGAlaArgGlnIuAlaIuArgAlaIuHisArgIuAlaIuHisIeuArg 265
612 TGCCAAATCGAAGACATGAATTCGCGCGCGCGCATCTGCGGTGTTGA 661
   ||::::::::::::::::::
266 LeuGlnArgATrArg.....GlyArgAlaAspGlyCysArgAlaI 279
662 GTGGCAGCAGATTCATTCATCGAGCGGT.....CGGC 696
   ::::::::::::::::::::
279 nGlyArgAlaArgAlaIuArgAlaIuArgIuGlnIuGlnIuProAla 296
697 GCGAATAAACCGTGTGACCATCAATTAATCAAGATGTAATTCATGG 746
296 rGlu.....AspArgAlaIuHisArg 304
747 CCG.....TT 751
305 ProArgProArgIuArgIuGlnIuAlaIuArgIuAspArgAlaIuG 321
752 TGTTTGCAGACGCGCTGTGAACCGGCGGCGATGTCCTAG.... 797
321 uAlaIuArgIuArgProAlaIuArgProArgIuGlnIuGlnIuProAla 338
798 .....TGTTCTCAATCAACAAACCGCGCTGTGCGCATTCG 836
338 lAspIeuProHisIuGlnIuProgluIuAlaIuArgIuAspArgI 354
837 TTTGGTGGAAGATTCGCAATTAATCGCGGCGCAATGTTGAGCAAG 886
355 GlnIuIuGln.....GlyAlaIuArgIuSerArg 365
887 ACAACCGCGATTCGCGTTCGCGTATTAAT.....CGCGCGATTA 930
365 gHisProHisArgIuAlaIuArgIuGlnIuGlnIuArgIuAspArg 382
931 CAAGCGCGCAGATTAATTTGGAGCGGATCAACAAATGAGATTCGTTAT 980
382 euArgArg.....ProArgATrArgArgIuGln 391
981 CGAAGAGCGCGCAAGAAAGCTGTGCGGTGGCGGCGGCGCGG 1030
392 ArgArgGlnProArgGlnIuGln.....GlyAlaHisIuGln 404
1031 ACAATTAATCATGACGCGGTACAAACCGCGCGCTTCTGAAAAACAA 1080
404 yGlnIuIuArgIuHisIuArgIuArgArgArgIuAlaIuProSerGln 420
1081 CTCTTCAAGTTCAACAGACCGGCTCAACGCGGCGGCGCGCGCATGTTG 1130
421 leuAlaIuArgIuGlnIuAlaIuArgIuAlaIuSerAlaIuHisIuAla 437
1131 GATTGGTATTCAGAGCGCGGTGATGCGCTTGGATATCTGCGCACCGCTG 1180
437 uTrpMetHisSerIuArgIuGln.....LeuArgIuAlaIuArgIu 452
1181 TTTTGCAGATTAATCGTGGCGATACGAC..... 1212
452 euCysArgAspSerValGlnSerHisIeuAspValThrTrpArgAla 468
1213 .....AGCGGCAAGCATTTGGTGTGGAATTCGAGCAAGA 1250
469 AspAlaSerIleSerArgATrThrTrpHisValMetAsn...ThrArg 484
```



```

631 ..... 631
210 CGCAGTAACACACGCGGATCTTTGTCTTCAACAGCATTGGC 161
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
632 .AlaLeuHisSerLeuProAsnGlyAlaLeuAspGlyLysAlaSer.... 646
160 CTTTGTGACGGCATCGCTTCTTGTGCTTTCATCGAGGGGCGCATACCG 111
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
647 .....LeuLeuSerSerProAlaAlaThrAlaLeuAspAsnGlyLeuThr 661
110 GCATATTCTTCGCCACGACGACCTGGTATGCGCGGCGC 66
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
662 SerArgSerThrLeuSerSerGlySerValSerLeuSerPro 676

seq_name: sp_invertebrate:Q9N3Y8

seq_documentation_block:
ID   Q9N3Y8          PRELIMINARY;      PRT; 2344 AA.
AC   Q9N3Y8;
DT   01-OCT-2000 (TREMBlrel. 15, Created)
DT   01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT   01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE   Y40C5A.3 PROTEIN.
GN   Y40C5A.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Pelodierinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN-BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018 (1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN-BRISTOL N2;
RA   Kallunki J., Randal J.;
RT   "The sequence of C. elegans cosmid Y40C5A.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN-BRISTOL N2;
RA   Waterston R.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR   EMBL; AC024772; AF60538.1; -.
DR   InterPro: IPR000194; ATPase_alpha_beta.
DR   InterPro: IPR002965; P-rich_extensn.
DR   PRINTS: PR01217; PRICEXTENS.
DR   PROSITE: PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
SQ   SEQUENCE 2344 AA; 245902 MW; 7D8696D48D0ADDBB CRC64;

alignment_scores:
Quality: 110.50      Length: 529
Ratio: 0.505        Gaps: 22
Percent Similarity: 41.399   Percent Identity: 20.038

alignment_block:
us-09-303-518d-125/rev x Q9N3Y8 ..
Align seg 1/1 to: Q9N3Y8 from: 1 to: 2344

1304 AGCGGCGCGTATTGTAATTCGCCGCGACAGACGTCGACAAACGCGAG 1255
    ||| ||| ||| ||| |||:|||||:|||||:|||||:|||||:|||||:
1192 SerProProAlaSerSerLeuSerAlaSerAlaLysLeuProLysSerG1 1208
1254 GTCTTCTTCGTCATTCACAGACCAACCAATGCGTGGCGTGTGCGTAT 1205
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1308 uSerGlyPheMetAsnAlaIleProPro.....S 1218

1204 CGCGGAGATTAAATCGCGCAAAAGCAGGGTGGCAGAGATATCAAGGCG 1155
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1218 erGluPheValAspSerSerSerGlySerSerGlySerPheGlySerGly 1234
1154 ATCAAGGCGCGTGTAGTACCAATCGGACCATGCGCGCGTGGCGG.... 1110
1235 LeuGln.....IleProThrThrAspSerSerThrProMetAs 1248
1109 .....CGTTGACGCGTG 1097
1248 ProSerMetGluAspThrAspValAspLeuThrValProThrThrAlaV 1265
1096 TGTTCGAACCTTGAAGAGTTGTTTTCAGAGAAATGGCCGAGGTTGTACGC 1047
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1265 alThrSerSerProValLeu..... 1271
1046 GTGATGAGATATTGTCCGCGTGGCGGCGACACCGACGACGCTTTT 997
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1272 ..... 1276
996 GCTGCGGCGCTTTCGATACGGAATCTGATTGTGTAAGCGTCCCAAT 947
    ||| ||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1276 uSerArgGlnSerLeuIleThrSerIleIleSerSerGlyLysProAla. 1292
946 AATCGTGGCGCGCTTGTGAATCGCGCGTTCGAATACCGAACCAGGAATC 897
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1293 ..ThrIleProAspThrIleMetAlaMetLysThrSerGlnSerSer 1308
896 ACAGCGGTGCTGTGTCACCAATTCGCCCGCAGTAAT..... 858
1309 SerLysPheSerLysSerThrAlaPheProThrLeuIleThrLysGlnTh 1325
857 .....TCGATACTTTCGACCCCAA..... 837
1325 rGluIleLysProThrThrValSerSerPheAlaProLysValProGluT 1342
836 .....ACGTAACGCAAGAGCGCGGTGGTGTGACTTGA 804
1342 hrLysIleAsnGlnProThrIleSerSerAsnIleGlyAspLeuSerLys 1358
803 GAACACCACTGAGGCAATCAGCGGCTCGGTTCAGACGGCTGTGCAAA 754
1359 ThrGluLysGlnSerThrIleProArgValGluValThrProSerAlaSe 1375
753 CAACGCGCAATGTGAATTACATCTGTGAATGATGTCACACGGTTT 704
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1375 rProGlnSerThrThrValThrGlu.....His.....I 1385
703 TATTGCGCGCGACCGCGCTCGCATGAATGAATGTCGTTG..... 666
1385 leAspAlaGlnThrAsnSerAlaThrIleProThrIleGlyThrProPro 1401
665 ...CCACTCAACCGCGCAGATGCGGCGCGCGCATTCATGCTTTCGAT 619
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1402 SerProLeuLysProLys.....Thrly 1409
618 GTTGCGACGATTTTCAGACGCGACGTCGCGCAGTGCCTTCAACAAT 569
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1409 sleuGlyLeu.....ThrSerHisProSerAlaIleProProT 1422
568 GGATTTTGGCGTTCGTCGAACGCGTC..... 543
1422 rPalAlaIleSerSer...LysThrLeuAlaProProValAlaProProThr 1437
543 ..... 543
1438 ValThrValProSerAsnIleAlaProSerThrThrGlyHisGlnSerG1 1454
542 .....AATCAACAGAGCGCGCGTGG..... 522
1454 nGlnThrArgProThrProThrThrHisArgProGlyIleThrProProL 1471

```



```

547 CGTTTGACCGAAGCAAAATCCATGTTGTAAGCAGCTGGCGAGACGT 596
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
994 .....ProHsthValGlnIlePheSerProIlePheLeuGlnProL 1008
597 GCCGCTGCAAAATGCTGCCAACATCGAAGAACATGAA.....TTGGC 639
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1008 ysArgLeuThrLeuIleProAspSerLysPheGlnLeuGlnValValGly 1024
640 GGCCGCGATCCCTGCG.....GCTTGAGTGGCAGCAGCAGCAT 674
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1025 GlYProGlnIleProThrProProLeuAspPheSerLeuAsnSerMetI 1041
675 TCATTTTCATCGAGCCG.....GTCCGCGCGA 700
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1041 eAlaSerIleGlnProAsnAlaLeuIleThrSerSerGlnLeuGlyTyr 1058
701 ATAAACGCTGGGACCATCAATTATCAAGAT..... 732
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1058 hAlaIleThrGlyThrValArgValGlyAspGlnHisValThrLeuAsp 1074
733 .....GTAAATTCATGCGCGCTGTTGTTGCACACA..... 762
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1075 ThrValValLeuArgValAlaSerLeuGlyGlyIleIleLeuSerAla 1091
763 .....GGCGCTGTGAAC..... 774
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1091 rSerArgLysValGlnThrGlyArgValAlaAsnLeuArgLeuArgGly 1108
775 .....ACCGAGCGCGTATGCGCCTGAGTGTCT..... 804
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1108 alIleAlaGlyAlaGlnAspGlnIleProPheAlaPheGlyGlyAlaIle 1124
805 .....CAAGTCACAAACCGCGCTCTTGGCTGAC 833
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1125 TyrProPheLysValThrTrpSerValSerAspProSerValLeuPhe 1141
834 CGTT.....TTGGTGGCAAGTATTCGCAAAATTAATCGCGCGAATTG 877
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1141 rThrHisProLeuGlyGlyAspValValGlnProThrAspAsnGlnPhe 1158
878 TTGACACAGACAAACCGCGTATTCGCTGATTAAGCGCGCAT 927
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1158 lAlleTrpPheAsnAlaIleArgGlySerValThrValLysAlaVal 1174
928 ACACAA.....GGCGGCGACGATTAATTGGACGCTACCAC.. 963
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1175 ValGlnLeuAsnGlnLysAlaArgLysHisPheThrGlyArgThrSer 1191
963 ..... 963
1191 rPheThrAlaGlnThrThrIleThrValGlnAspGlyLeuSerLeuVal 1208
964 .....AATCAGATTCCTGTTATCGAAGAGCGCGAGC 996
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1208 lProGlnLeuAspIleAsnThrValArgValAlaProAsnSerGlnLeu 1224
997 AAAGAGCTGTGCGCTGGGTTGCGCGCAG.....CCGACACAAATA 1037
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1225 LysMetValThrAlaTrpSerGlnAlaSerPheSerValProSerAsp 1241
1038 CTCC.....ATCAGCGGTACAAACCTCGGCATTCCTGAAACAAAC 1081
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1241 eSerSerArgIleValIleSerAlaAspGlyHisLeuIleThrAsnGly 1257
1082 TCCTCAAGTTCAACACAGCC.....GTCAACGCGGCG 1113
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1258 .....LysGlnGlySerAlaAlaIleThrValArgAsnValAsnSerPro 1272
1114 GACGCG.....GCCATGTCGCGATGTTGTTACTTACGAGCGCGGATGCC 1157
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1273 AspAsnGlnThrValLeuIleProVal.....ThrValSerArgValAla 1288

```

```

1158 CTTGATATCTGCGCCACCGCTTTGGCG 1188
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1288 rLeuAspValHisProThrIleGlnLeuLys 1298

```

seq_name: sp_bacteria:Q9ZA63

```

seq_documentation_block:
ID      Q9ZA63      PRELIMINARY;      PRT;      270 AA.
AC      Q9ZA63;
DT      01-MAY-1999 (TrEMBLrel. 10, Created)
DT      01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT      01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE      HYPOHETICAL 30.2 KDA PROTEIN.
OS      Neisseria meningitidis.
OC      Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX      NCBI_TaxID=487;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=B1940;
RX      MEDLINE=99005251; PubMed=9790590;
RA      Claus H., Frosch M., Vogel U.;
RT      "Identification of a hot-spot for transformation of Neisseria
RT      meningitidis by shuttle mutagenesis using signature-tagged
RT      transposons."
RL      Mol. Gen. Genet. 259:363-371(1998).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=B1940;
RA      Claus H.;
RL      Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF000250; CA03966.1; -.
KW      Hypothetical protein.
SQ      SEQUENCE 270 AA; 30192 MW; 612DE74D02BA09AA CRC64;

```

alignment_scores:

	Quality:	109.50	Length:	382
	Ratio:	0.720	Gaps:	25
	Percent Similarity:	39.791	Percent Identity:	26.178

alignment_block:

US-09-303-518d-125 x Q9ZA63 ..

```

Align seg 1/1 to: Q9ZA63 from: 1 to: 270
32 CCATCGCGGCGAGCCGAGCAAGCCGTTTACG..... 64
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
22 sProProValSerAlaProThrAlaLysProValSerGlySerLys.... 37
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
96 TGGCGAAGATATGCGGATATGCGCCCTCGATGAAAGTCAAGGAGCG 145
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
38 .....LysProAsnSerMetSerProLysAla 46
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
146 ATGCGGTCA...AAAAGCCAAATGCTGTTGAAAGCAAAAGCAATCCG 192
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
47 SerSerSerAlaLysAsnAlaLysGlyCysLeuLysProLysThrIle.. 62
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
193 GCGGTGG.....TGTACTCGCGCGGCTTC 218
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
63 .....TrpGlnAlaArgLysAsnProTyrSerThrIleCysProArgLeuP 78
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
219 AGCGAAATTCGCGCGGATTCACCTGCGCGAAGAGCGGTACTTCAGTCA 268
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
78 hArgMetSerAsnSerPheThrValSerAlaArgAlaProSerAlaArg 94
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
269 TCGTGATTCGCGTTGAAGCAAGCAAGCAAAATCGATTGACCGTACGCA 318
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
95 AsnArgPheProVal.....ThrLysSerProAlaSerSerThrAl 108

```



```

163 LeuMetAlaGluLysLeuGluGluIlePheSerGlyPheGlnIleLeu 179
520 .....TTCAACGCGCGCGTGTGATTTAGCCGTTTGACCG 556
179 PylsIlePheLysPheLysGluIleIleIleIleSerAsnLysAspL 196
557 AA.....CGCAAAATCCATGTTGTAGAGCAGCTGGCGCA 591
196 YslLeuLysLysGluPheGluLysLeuSerIlePheLysAsnAlaGlySle 212
592 GAGCGGCGGCT...GAAATGCTGCCACATGAAACACATGAATTCGG 638
213 LysIleLysSerLeuGluAsnAlaLysProLysThrAsnHisGluMet... 228
639 CGGCGCGCATCTGCGCGTTGAGTGCAACGCAATTCATTCATC... 684
229 .....IleMetHisPheLeuLys... 717
685 .....GAGCGCGTGGCGCGCAATTAACCGTGTGACCG 717
235 snAsnLysAsnThrLysAspAspIleAsnProHisAsnIleLeuLeu 251
718 ATCAATTATCAAGATGTAATTACCATTTGCGGTTGTTGCAACAGCGCG 767
252 AlaAsnIleGluAspLeuLysAsnAlaAsnLeuValIleLysAsnAsnAs 268
768 TCTGAACACCGAGCGCGTATTGCGCTAGTGTTCTCAAGTCAACACAAAC 817
268 nProLysLysGluLysPheValAlaIleAsnGlyAsnLysLysIleLys 285
818 CGGCGCTCTTGGTACCGTTTGGTGGCGAAAGTATCCAAATTAATCTCG 867
285 erArgIleLeuLysValIleLysIleGlyThrSerPheSerGlnLeu...Ile 300
868 GCGGAATTGGTTGACACA 885
301 AsnGluLysIleAspThr 306

```

seq_name: sp_invertebrate:Q22579

seq_documentation_block:

ID Q22579 PRELIMINARY; PRT; 1844 AA.

AC Q22579;

DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE HYPOHETICAL 187.1 KDA PROTEIN.

GN T19D12.1

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Favallo A.;

RT "The sequence of C. elegans cosmid T19D12.1";

RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Waterston R.;

RT "Direct Submission";

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; U41263; AAC24428.2; .

KW Hypothetical protein.
SQ SEQUENCE 1844 AA; 18713 MW; 6151A72777517F5 CRC64;

alignment_scores:
Quality: 109.50 Length: 317
Ratio: 0.788 Gaps: 14
Percent Similarity: 43.849 Percent Identity: 24.606

alignment_block:

US-09-303-518d-125/rev x Q22579

Align seg 1/1 to: Q22579 from: 1 to: 1844

```

935 CTTGTGTAATTCGCGCGTTCATTAACCAACCGGAATACCGCGTTC 886
310 ProValValThrThrSerThrSerThrGlnGlyLeuSerThrThrAl 326
885 TGTGTCAACCAATTCGCGCGCATTAATTCATTAATTCATTAATTC 858
326 AlaAlaThrProSerSerSerValIleProThrThrThrGlnThrG 857
857 .....TGCGATCTTTCGACCCCAACCGTACGCAAGAGCGCGT 816
343 InArgProThrSerThrGlyLeuProSerThrValSerThrSerGln 359
815 TTGTGACTTGTAGAACACCTAGGCAATACGCGCGTGTGTTCAAGAC 766
360 ThrSerSerThrSerProIleProSerThrThrGlnThrSerSerAl 376
765 GCCTGTGTCAACCAACCGCAATGTTAATTAATTCATTAATTAATG 716
376 apo.....SerT 379
715 TCACACGCTTTTATTCGCGCGACCGCGTGTGATGAATGATGCGTG 666
379 hrLysThrSerAsnThrProSerProThrThrLeuLeuThrSer 395
665 CCACTCMAACCGCGCAGATGCGCGCGCGCAATTCATGTTTCATGTT 616
396 ThrIleAlaProSerThrGlnGlyValProThrSerSerSer..... 410
615 GGCACATTTTCAGACGCGCAGCTGCGCGCGCGCTTCAACACATGA 566
411 .....SerSerProAsnSerThrProThr..... 418
565 TTTTGGCTTGGTCAACCGCGCTCAATACCAACAGCGCGCTTGAATCC 516
419 .....ThrThrIleThrProGlyAlaProSer 427
515 TCGGCGGCTTCTTGATTAATGACCGGTAGGCGCGCAGCGATGTTGT 466
428 Ser.....ThrLeuGlySerSerSerSerThrIleVal 438
465 GTTCATGCGATTGACAGATGCGCAACGCGCTGCGATGAGCGAGGAA 416
438 lserThrThrIleThrPro...SerThrProLysValSerThrIleThrL 454
415 TTTTGTGACGAGCGCGGTGCGCAGCTGCAACACCGGATGTGATC 366
454 eu.....SerGlnSerProThrPro..... 460
365 AGTTGCGGCGCACTTCTGCGCGCTTAAGTTGCGAGCGTTGAGTGC 316
461 .....ThrSerThrProLeuValValSerSerSer..... 471
315 GTAGCGTTCAACATTCATTTGCTGCTTCAACGCGCAATACGACTG 266
472 .....SerGlySerSerSerThrValValThrSerThrIleThrPro. 485
265 ACTGAATGACGCGCTTTCGCGCAGGTGAATGCGGCGCATTTTGCCTGA 216
486 .....SerThrGlnGly.....ValProThr 492

```

215 GCCGGCAGTAACACACCGCGATCTTTTGTCTTCAACACAC 166
 493 SerThrSerAsnGlnProThrProser.....ThrsSerAsnProth 506
 165 TTGGCCT.....TTTTGACGGCATCGCCTCTTGCATTTCACGAGG 122
 506 rThrProLysSerThrValThrAlaSerProserThrThrGlyAlaThrS 523
 121 GCGGCATA...CCGGCATTTCTTCCGACACACGCGACTTCGGTATG 75
 523 erThrAlaSerProserThrIleThrSerSerAlaProThrSerGlnSer 539
 74 GCGGCGCCTGCTAAACGGCTTGTCCGGTCTGCCCGCATGGCAGGTT 25
 540 HisSerProserSerThrMetThrSerThrValProValThrSerThrPh 556
 24 T 24
 556 e 556

seq_name: sp_invertebrate:P91365

seq_documentation_block:
 ID P91365 PRELIMINARY; PRT: 2232 AA.

AC P91365;
 DT 01-MAY-1997 (TREMBlrel. 03, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE K06A9.1 PROTEIN.
 GN K06A9.1
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pelodermidae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Geisel C., Gattung S.;
 RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS (A AND B) MAY BE PRODUCED BY
 CC ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE SHOWN IS THAT
 CC OF THE A FORM.
 DR EMBL: U080846; AAC70889.1; -;
 DR EMBL: U080846; AAC70890.1; -;
 KW Alternative splicing
 FT VARSPLIC 842 866
 FT VARSPLIC 870 1051
 FT VASSPABSTSONPMPNPNSSGSSMI -> IATTSAPKPSVT
 FT CLEMTD (IN ISOFORM B).
 FT PYPSSSTPESSTTPSPGSPETTLITSTSPSSSTTIGST
 FT OGSTPGLSTSEMTSGSTQTPGSTSTVTPSTVQSDST
 FT SSGSTVVGSTEGSSSPSTSONTPSTSGSSMTQTPQ
 FT SSOSTPVESTSGATSSSGSTGLTISPSPPSTIGS
 FT SSGSTSPVSTISOGSTE -> KEIDOTAIWTKYENFAL
 FT LVASKINNESITGYIDNPGYSAGINDHOQYPTDYNKIKS
 FT VPPIDGTDDBDLKDYDKSLATADMTPPVADOTCMIFI
 FT SAAPDEYGGTTIKSTYEFVGVLVGAKSPDGLSTDK
 FT NIVITNNTNMDRDAASAVVSKLELLPLTA (IN ISOFORM
 FT B).
 SO SEQUENCE 2232 AA; 213840 MW; 08D69FA638E14C8 CRC64;

alignment_scores:
 Quality: 109.50 Length: 458
 Ratio: 0.495 Gaps: 18
 Percent Similarity: 48.253 Percent Identity: 22.926

alignment_block:

US-09-303-518d-125/rev x P91365

Align seg 1/1 to: P91365 from: 1 to: 2232

1340 CCTCTCTTCATAGTGTTCGACACTTGGCAGACACGCGCGATATC 1291
 |||: ||| ||||| ||| |||

1425 ProThrGlySerThrGlySerThrLeuSer.....se 1437
 1290 GTATTGGCCGGGACGAAAGCTGCACAAAGGAGTCTTCTGCTCA 1241
 1437 rThrIleSerGlySerThrGln...HisThrThrMetSerLysAlaSerS 1453
 1240 ATTCGACCAACCCAAATGCCCTGCGCGCTGCGATGCGGACGATTAA 1191
 1453 ergly.....SerThrSerProserThrAsn 1461
 1190 TCGCGCAAAAGCAGGTGGCGAGATATCCAAAGGCATCACGCGCTGCT 1141
 1462 SerGln.....ThrGlySerThr 1467
 1140 ACTACCAATCGGACCATGGCGGTCGCGCGCTGACGCGCTGTGTGA 1091
 1467 rValThrMetGlySerSerThrSerGlyValSerThr..... 1480
 1090 ACTTGAAAGATTGTTTTCAGAAATGCCGAGGTTGTACGCGTATG 1041
 1480 1480
 1040 GAGTATTGTCCGCGTGGCGGACACCGGAAACACTCTTGTGCTGC 991
 1481SerSerAlaSerThrGlnProGlnMetSerThrSerG 1494
 990 GCCTTCTGATACGAATCTGATGTGTTAGCTCCCAATATTCGT 941
 1494 nglySerSerAlaGlySer.....ThrV 1502
 940 GCGCGCTTGTGTAATCGCGCCG.....TTCATACCGAACCGGAATC 897
 1502 alAlaSerSerThrAlaSerProAlaAlaSerSerThrAlaProSer 1518
 896 ACAGCGGTGCTGTGTCAACCAATTCGCCCGCAGTA.....ATTG 856
 1519 ThrGlyThrMetSerSerThrSerSerGlyThrValGlySerThrIle 1535
 855 CGATCTTTCGACACCAAAAGGTACGCAAGAGGCGGCTTGTGACT 806
 1535 rGlySerSerThrThrAlaSerAlaSerSerGlnThrGly...SerThrV 1551
 805 GAGAACCACTAGGCAATACGCGCTGCGTTCGAGCGGCTGTGCA 756
 1551 alThrMetGlySerSerSerThrSerGlyValSerThrSerSerAlaSer 1567
 755 AACAAAGGCCAATGTAATTAATTAATTAATTAATGTC...CAAC 709
 1568 SerThrGlnProGlnMetSerThrSerGlnGlySerSerAlaGlySer 1584
 708 GGTTTTATTCGGCGCGCGCGCTCGATGAATGAATGTCGTCGCACTCA 659
 1564 rValAlaSerSerThrAlaGlyLeuValSerThrSerThrValProSerS 1601
 658 AACCGGAGATGCGGCGCGCAATTCATGCTTTCATGTTGGCAGCA 609
 1601 erThrGlyThrMetGlySerThrSerSerGlyThrValGlySerThrIle 1617
 608 TTTTCAGACGCGACGCTGCGCGCGCTTTCATGAACATGATTTTG.. 561
 1618 SerGlySerSerThrThrAlaSerAlaSerSerGlnThrGlySerThr 1634
 560CGTTCGTCMAACGCGCTCAATA 539
 1634 lThrMetLysSerSerThrSerGlyValSerThrSerSerAlaSerS 1651
 538 CCACAGCGCGCGTTTGAATCTTCG...GCGGCTTTCATGAATAGAC 492
 1651 erThrGlnProGlnMetSerThrSerGlnGlySerSerAlaGlySerThr 1667
 491 GTAGGCTGGCAGCGCAGGATGTGTGTCATTCATTCAGCAAGATGCG 442
 1668 ValAlaSerSerThrThrGlyLeuValSer.....ThrSerThrVa 1681

```

441 GAACGGCTGGCATGACGAGCATTTTGTGAACGAGGCGTGGCA 392
   : : : : : : : : : : : : : : : : : : : : : : : :
1681 lProserSerThrGlyThrMetGlySerThrSerGlyThrValGlyS 1698
   : : : : : : : : : : : : : : : : : : : : : : : :
391 GCGCAGTCCACAAACCGGATTGGATC.....AGTTGGCGGCACCTTCT 348
   : : : : : : : : : : : : : : : : : : : : : : : :
1698 erThrIleSerGluSerThrAlaIleSerAlaSerSerGlnThrGly 1714
   : : : : : : : : : : : : : : : : : : : : : : : :
347 TCGCGCGTTAGTT.....GCCAGCGCTTCAGTGGTGGTCAAA 304
   : : : : : : : : : : : : : : : : : : : : : : : :
1715 SerThrValThrMetGlySerSerThrSerGlyValSerThrSerSe 1731
   : : : : : : : : : : : : : : : : : : : : : : : :
303 CTGATTGCTGCTTGGCT.....TCACGGCGCA 275
   : : : : : : : : : : : : : : : : : : : : : : : :
1731 rAlaSerSerGlyGlnProGlnMetSerThrSerGlnGlySerSerAla 1748
   : : : : : : : : : : : : : : : : : : : : : : : :
274 TCAGACTGACTGACGCGCTTTGCCGACGCGTGAATCGCGCGATT 225
   : : : : : : : : : : : : : : : : : : : : : : : :
1748 lSerThrValValSerSerThrAlaSerProAlaIleSerSerThrAla 1764
   : : : : : : : : : : : : : : : : : : : : : : : :
224 TTGCGTGAACCGCGCGAGTAAACACACGCGCC.....CGATT 187
   : : : : : : : : : : : : : : : : : : : : : : : :
1765 ProSerSerThrGlyThrMetSerThrSerSerGlyThrValGlySe 1781
   : : : : : : : : : : : : : : : : : : : : : : : :
186 CTTTGTGCTTCAACAGCATTGGCTTTTGTGACGCGATCGCTTCT 137
   : : : : : : : : : : : : : : : : : : : : : : : :
1781 rThrMetSerGlnSerSerThrAlaIleSerThrSerThrGlyS 1798
   : : : : : : : : : : : : : : : : : : : : : : : :
136 TGACTTTCATCGAGGCGCATACGCGCATATTCTCCGCAAGCAGCGG 87
   : : : : : : : : : : : : : : : : : : : : : : : :
1798 erThrValThrLeuGly.....SerSerSerThrSerSerGln 1811
   : : : : : : : : : : : : : : : : : : : : : : : :
86 ACTTCGCTAATGCGCGCGCGCTCG 63
   : : : : : : : : : : : : : : : : : : : : : : : :
1812 MetSerThrSerGlnGlySerSer 1819
   : : : : : : : : : : : : : : : : : : : : : : : :
seq_name: sp_bacteria:Q9KXT0

seq_documentation_block:
ID Q9KXT0 PRELIMINARY; PRT; 960 AA.
AC Q9KXT0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE PUTATIVE SUGAR HYDROLASE.
GN SCSH4_09C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Murphy L., Harris D.;
RA Submitted (May-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RA Submitted (May-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinsahl H., Hopwood D.A.;
RA "A set of ordered cosmids and a detailed genetic and physical map for
RA the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RA MCL. Microbiol. 21:77-96(1996).
DR EMBL: AL35913; CAB91121.1;
DR InterPro: IPR001764; Glyco_hydro_3.
DR InterPro: IPR002772; Glyco_hydro_3C.

```

```

DR Pfam: PF00933; Glyco_hydro_3; 1.
DR Pfam: PF01915; Glyco_hydro_3_C; 1.
DR PRINTS: PR00133; GLHIDLASE3.
KW Hydrolase.
SQ SEQUENCE 960 AA; 102056 MW; 5332C02428BF39A2 CRC64;

```

```

alignment_scores:
  Quality: 109.00      Length: 470
  Ratio: 0.522        Gaps: 27
  Percent Similarity: 44.468      Percent Identity: 23.191
alignment_block:
US-09-303-518D-125 x Q9KXT0

```

Align seg 1/1 to: Q9KXT0 from: 1 to: 960

```

46 CCGGAGCAGACCGCTTACGAGCGCGCCGATTCACGAAGTCGCTGCT 95
   : : : : : : : : : : : : : : : : : : : : : : : :
319 ProGluAsnAspProTyrSerGlyAlaIleAspThr..... 330
   : : : : : : : : : : : : : : : : : : : : : : : :
96 TGCGAAGATATGCGCGTATCGCGCCCGCATGAAATCAAGCAGAGCG 145
   : : : : : : : : : : : : : : : : : : : : : : : :
331 .GlyAlaAspThrAsp.....ThrProGluHisArgAlaLeuAlaArg 345
   : : : : : : : : : : : : : : : : : : : : : : : :
146 ATGCGGCTCAAAAAAGCGCAAGTCTGTTGAAGACAAAGAAATCGCGG 195
   : : : : : : : : : : : : : : : : : : : : : : : :
345 sPAlaIleGluGlnAlaValValLeu.....LysAsnAspGly 358
   : : : : : : : : : : : : : : : : : : : : : : : :
196 GTGGTGTTCACGCGCGCGCTTCAGGCAAAATCGCGCGATTCCGCTGG 245
   : : : : : : : : : : : : : : : : : : : : : : : :
359 ValLeuProLeuAlaPro..... 364
   : : : : : : : : : : : : : : : : : : : : : : : :
246 CCAAAAGCGCGTACTTCACTGATCGATGATGCGGTTGAAGCAGAGCG 295
   : : : : : : : : : : : : : : : : : : : : : : : :
365 .GluThrArgValAlaValAlaGlyLeuLeuAla.....AspG 377
   : : : : : : : : : : : : : : : : : : : : : : : :
296 AAATGAGTTTGAACGCTACGACCTGAGAGCGCTGGCAAACTTAAGCGC 345
   : : : : : : : : : : : : : : : : : : : : : : : :
377 LuCysLeuLeuAspTrpTyr.....SerGly 385
   : : : : : : : : : : : : : : : : : : : : : : : :
346 GAAGAGAGCGCGCGCAACCTGATCCAAATCGGTTTGTGG..... 384
   : : : : : : : : : : : : : : : : : : : : : : : :
386 ThrLeuIleHisArgSerThrProLeuGlnGlyLeuTyrGluArgPheG 402
   : : : : : : : : : : : : : : : : : : : : : : : :
385 .....ACTGCGCTCGCA 397
   : : : : : : : : : : : : : : : : : : : : : : : :
402 yAlaAspArgValSerPheAlaGluGlyValAspArgValArgLeuArgT 419
   : : : : : : : : : : : : : : : : : : : : : : : :
398 CC.....CGTCGTTGACGAAATTCCTGCGCGATGCG..... 432
   : : : : : : : : : : : : : : : : : : : : : : : :
419 hrAlaAspGlyArgPheLeuHisValLeuProAlaAspSerAlaSerAla 435
   : : : : : : : : : : : : : : : : : : : : : : : :
433 .....GAGCGCTGCGCATCTTCGTCA 455
   : : : : : : : : : : : : : : : : : : : : : : : :
436 GluAlaProGlyThrGluGlyAlaLeuAspPro...AlaLeuLeuAlaG 451
   : : : : : : : : : : : : : : : : : : : : : : : :
456 TGCATGACACCAATACCGCTGGTCGCGACCTACGATCATTAAGAAG 505
   : : : : : : : : : : : : : : : : : : : : : : : :
451 yArgThrAspLeuProProLeuThrThrAsp...AlaValGlyThrGluL 467
   : : : : : : : : : : : : : : : : : : : : : : : :
506 AAGCGCGGAGATTTCAAACGCGCGCTGTGATTATGAGCGTTTGACC 555
   : : : : : : : : : : : : : : : : : : : : : : : :
467 euAlaLeuIleAspTrpGlyGluGlyValLeuThrLeuArgAlaSerAsp 483
   : : : : : : : : : : : : : : : : : : : : : : : :
556 GAACGCAAAATTCATGTTGTAG.....GCAGCTGGCGCAGA 593
   : : : : : : : : : : : : : : : : : : : : : : : :
484 GlYArgTyrLeuSerValAlaGluAspGlyPheValArgAlaSerAlaAs 500
   : : : : : : : : : : : : : : : : : : : : : : : :
594 CGTGGCGTCT.....GAAATGCTGCCAATCAATGCAACAGCATG 631
   : : : : : : : : : : : : : : : : : : : : : : : :
500 pGlnProGlyGlyTyrPheValValGlnGlnThrPheArgLeuGlnProHis 517

```

```

632 AATTGGGGGGGGGGGATCTCCGGTTTGAATGCGACGACATTCATTTG 681
    |||
517 SPASpLy.....HisLeu 521
682 ATCGAGCGCGTGGCGCGAATAAAACCGTGTGAC..... 717
    |||
522 LeuArGhIstHrGlyThrGlyArgProValGlnValAlaAlaAspGlyVa 538
    |||
718 ...ATCAATTATCAAGTAAATTAATACATGGCGGTTTGTTCACAG 763
    |||
538 llysvAlaAlaAlaProAspAlaAlaAlaGlyAlaAlaAlaAlaGlyAla 555
    |||
764 GCGGTCTGAACACGACGCGCGTGTATGCTAGTGGTTCACAGTCAAC 813
    |||
555 lYserAlaAspAlaGlu...ValPheGlnLeuValAlaGluArgGly 570
    |||
814 AAACCGCGCTCTTGGCGTACCGTGTGGTGGCGAAGTATGCAAAATTC 863
    |||
571 GluAspAlaValThrArgValAlaAlaGlyAlaAspAlaValAlaValVa 587
    |||
864 TGGGGCGAATTGTTGACACAGACACGCGGTGATTCGGTTCGGTAT 913
    |||
587 lAlaGly.....AsnAspProHisI 594
    |||
914 TGAACGCGCGGATTCACACAGCGCGCGATTAATTGGAGCGTACACAG 963
    |||
594 lAsnGlyArgGluThr.Glu.....AspArgThrTh 604
    |||
964 AATCAGATTTCGTTATCGAAGAGCGCGACAAAGAGTGTGGCGTG 1013
    |||
604 rLeuAlaGluProAlaGlnGlnGlnArgLeu..... 614
    |||
1014 GGTGGCGCGCGACGCGACAAATATCTCATTCACGCGTA..... 1051
    |||
615 ..LeuArgAlaAlaArgAlaAlaAsnProAlaThrValLeuAlaLeuVal 630
    |||
1052 .....CAACCTCGCGCATTCCTCTGAAAAACAACTCTTCAAGTTCAAC 1095
    |||
631 SerAlaTyPProTyAlaVal.....GlyValGluTh 641
    |||
1096 ACAGCGCTCAACGCGCGCGACCGCGCATGTCGCGATTCGTAATTCGA 1145
    |||
641 rLeuProAlaValLeuThrPthrAla.HisGly..... 651
    |||
1146 GCGCGGTATGCCCTTGATATCTGCCACCGCTTTCGGCGATTTAA 1195
    |||
652 .....GlyGlnAlaAlaGlyThrAlaLeuAlaArg.Ileu 663
    |||
1196 TCGTGGCGGATACGACGCGCGCGCATTTG...GGTTGCTTGAATTG 1242
    |||
663 euAlaGlyAspValSerProAlaGlyArgLeuProGlnThrTrpTyrSer 679
    |||
1243 GACGAAGAAGACCTCGCTTGTGCGAGCTTGTCCCGCGCAATTCGA 1292
    |||
680 AspAspAlaAspLeu.....ProGlyLeuLeuAs 689
    |||
1293 ATAC 1296
689 PTyr 690

```

```

seq_name: sp_human:09UHA8
seq_documentation_block:
ID 09UHA8 PRELIMINARY; PRT; 2296 AA.
AC 09UHA8;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE SPLICING COACTIVATOR SUBUNIT SRM300.
GN SRM300.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=20132238; PubMed=10668804;
RA Blencowe B.J., Baurten G., Eldridge A.G., Issner R., Nickerson J.A.,
RA Rosolina E., Sharp P.A.;
RT "The SRM160/300 splicing coactivator subunits.";
RL RNA 6:111-120(2000).
DR EMBL; AF201422; AAF21439.1;
DR InterPro; IPR002965; P_1rich_extensn.
DR PRINTS; PRO1217; PRICHEXTNSN.
SQ SEQUENCE 2296 AA; 251964 MW; 17C0BD4FA10A9CF9 CRC64;

alignment_scores:
    Quality: 111.50      Length: 435
    Ratio: 0.578        Gaps: 20
Percent similarity: 44.368      Percent identity: 23.678

alignment_block:
US-09-303-518D-125 x 09UHA8 ..

Align seg 1/1 to: 09UHA8 from: 1 to: 2296

18 AGGTCTAAACCTGCCATCGCGCGGCGACGACGACCGCTTACGACG 67
    |||
483 ArgSerArgSerProAlaThrAlaLysArgGlyArgSerArgSerArgTh 499
    |||
68 GCCC.....GGCCATTACGAGATCGC...CTTCTTGGCGAGAA 105
    |||
499 rProThrLysArgGlyHisSerArgSerArgSerProGlnArgArg. 515
    |||
106 TATGCCGCTATGCGGCC.....CTCGATGAAGTCAAGAGAGCGGA 146
    |||
516 ..SerArgSerAlaGlnArgTrpGlyArgSerArgSerProGlnArgArg 531
    |||
147 TGCCTGCAAAAAAGCCAGT..... 167
    |||
532 GlyArgSerArgSerProGlnArgProGlyTyrPserArgSerArgAsnTh 548
    |||
168 .....GCTGTTGAAGACAAAAAAGATCCGGCG 195
    |||
548 rGlnAlaArgGlyArgSerArgSerAlaArgGlyArgSerHisSerA 565
    |||
196 GTGGTGTTCATCGCGCGCTTCAGCAAAATCGCGCGATTCACCGTGG 245
    |||
565 rGserProAlaThrArgGly.....ArgSerArgSerArgThrProAla 579
    |||
246 CGAAAGCGGCTACT.....TCAGTCAG 268
    |||
580 ArgArgGlyArgSerArgSerArgSerArgThrProAlaArgArgSerArgSe 596
    |||
269 TCGTGAATTCGCGTTGAAGGACGACGAAATCGATTTCAGCTTCACGGA 318
    |||
596 rArgThrProThrArgArgArgSerArgSerArgSerArgThrProAlaArgArg 613
    |||
319 CCTAAGCGCTGGCAACTTAAGCGCGGAGAGTGGCGCGCAACTCTGAT 368
    |||
613 lYArgSerArgSerArgThrProAlaArgArgArgSerArgThrArgSer 629
    |||
369 CCAATCGGTTTGTGAGTGGCGTGGCGACCGCTC....CGTTGACGAAA 414
    |||
630 ProValArgArgArgSerArgSerArgSerArgSerProAlaArgArgSerArg 646
    |||
415 ATTCCTGCCGTCATGCCGAGCGCTTGGCATCTTGTGCAATGCGATGGA 464
    |||
646 gSerArgSerArgThrProAlaAlaArgGlyArgSerArgSerArg...T 662
    |||
465 CACCAATCGCTGGCTGGCGACCGCTACGATCATTAATCAAGAACCGCGG 514
    |||
662 hrProAlaArg.....ArgGlyArgSerArgSerArgThrProAla 675
    |||
515 AGGATTTCAAAAGCGCGCTGTGGTATTCAGCCGTTTGACGGAACGGA 564

```

```

676 ArgArgSerGly.Arg.....SerArgSerArgT 665
565 ATTCATGTTTGTAGCGACGCTGGCGGACGAGTCCCTCGTAATAATCTCG 614
665 hrpP01A1AArgArgGlyArgSerArgSerArgTThrProAlaArgGlyArg 701
615 CAACATCGAACAACATGATTCGGCGCCGCGCATCCGCGGCTTTAGAG 664
702 SerArgSerArgSerLeuValAlaArgGlyArgSer..... 713
665 GCAGCGACATTCATTCATTCGACCGGT.....C 693
714 .HisSerArgThrProGlnArgArgGlyArgSerGlySerSerSerGluA 730
694 GCGCGCAATAAACCCTGTGACATTCATTCAGATGATTAATACAT 743
730 rglYsAnLysSerArgThrSerGlnArgArgSerArgSerSerSer 746
744 TGCGCGTTTGTTCACACAGCGCGCTGACACCGACGCGCTGAT..... 788
747 ProGluMetLysSerArgIleSerSerArgArgSerArgSerLeuSe 763
763 rSerProArgSerLysAlaLysSerArgLeuSerLeuArgArgSerLeuS 780
789 .....TGCCCTAGGTGGTGTCTCAAGTCAACAACCGCGCT 824
780 enGlySerSerProCysProLys.....GlnLysSerGlnThrProPro 794
825 CTTCGCTACCGTTTGGTGGCAAGATACGCAATTAATCTCGCGGCGAAT 874
794 ..... 794
875 TGCTGACACAGACAAACCGGTGATTCGGTTCGATGACGCGCG 924
794 ..... 794
925 ATTACACAGCGCGCAGATTATTGGAGCGTACACAAATCAGATTTC 974
795 .....ArgArgSerArgSerGlySerSerGlnProLysAlaLysSer 808
975 CGTTATCGAAGAGCGCGCAGCAAGAGCTGTGCGCGGTGGTGGCGCG 1024
809 ArgThrProProArg.ArgSer.ArgSerSerSerProProProLys 824
1025 AGCCGCGAACAATATCTCCATCAACGCGTACACCGTTCGCGCAAT 1074
825 GlnLysSerLysThrProSerArg.....G1 833
1075 AACAACTCTTCAGTTCAACACAGCGGTACACGCGCGCGACCGCGCA 1123
833 nSerIleSerSerSerProHisProLysValLysSerGlyThrPro 849

seq_name: sp_bacteria:Q93RW3

seq_documentation_block:
ID Q93RW3 PRELIMINARY: PRT: 406 AA.
AC Q93RW3;
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE HYPOTHEtical 44.7 KDA PROTEIN.
GN SC139.10C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Warren T., Harris D.;

```

```

RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=9700351; PubMed=8843436;
RA Redenbach M., Kiser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinsht H., Hopwood D.A.;
RT "A set of ordered cosmid and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL MOL. Microbiol. 21:77-96(1996).
DR EMBL: AL591322; CAC38798.1;
KW Hypothetical protein.
SQ SEQUENCE 406 AA; 44724 MW; E64CFBBDCTA632F CRC64;

```

```

alignment_scores:
Quality: 108.00 Length: 410
Ratio: 0.632 Gaps: 21
Percent Similarity: 41.707 Percent Identity: 22.683

```

alignment_block:

US-09-303-518d-125 x Q93RW3 ..

Align seg 1/1 to: Q93RW3 from: 1 to: 406

```

102 AGAATATGCGGGATGATCGCCCTCGATGAAGTCAAGAGCGCATCGC 151
33 ArgValAlGlyLeuHisAlaProLeuLysPsn...ArgGlyThrGlnCysAr 48
152 TCAAAAAAGC..... 161
48 glnAlaAlaLeuHisGlyProAlaArgArgValAlaThrValAlaVala 65
162 ..CCAAAGTCTGTTGAAGACAAAGATCGGCGCTGTGTTACTGC 209
65 sPrProAspHisLeuAlaArgGlnArgAlaAlaAlaArg..... 77
210 GCCGCGCTTCAGGCAAAATCGCCGC.....GATTCACCGTG 244
78 .....ArgGlnArgGlnArgValAlaGlyLeuGlnGlnGlyThrGln 91
245 GCGAAAGCGCGTACTTACGACTGCTGATTGCGCTTGAAGCAAGCAG 294
91 sGlnArgGlyValArgValAlaArgValAlaArgValAlaGlnCysArgHis 108
295 GAATCGAGTTTGAACGCTAGCGCACTGAAGCGCTGGCAAACTTAAGCG 344
108 rglArgArgValAlaProArgLeuValHisValAlaArgLeuGlnAlaAlaGlnArg 124
345 CGAAGAGTGGCGCGCAACCTGATCCAAATCGGTTGTGTGATGCGCTGC 394
125 GlnArgAlaGlyPro..... 129
395 GCACCGCGCGCTTACGCAAAATTCCTGCGCTGATCGCGACCGCTGCGC 444
130 .....GlyValArgArgArgGlyValArgValAlaProArgH 140
445 ATCTTCGTCAATGAGATGACACCAATTCGCTGCGTGGCGCA..... 485
140 is.....HisArgAlaGlyAlaLeuArgLeuAspAla 150
486 .....CCCTACGCTATTATCAAGAGCGCGGAGATT 520
151 LeuAspAlaGlyLeuGlnArgGlnHisLeuGlnLeuArgArgValAla 167
521 TCAAGCGCGCTGTGTTGATTCAG.....CCGTTTGACCGAA 558
167 glnArgLeuAlaProAlaValAlaProAlaLeuLeuProValAlaGlnGln 184

```

```

559 CGCAAAATTCATGTTTGTAAAGCAGCTGCGAGACGTGCCGTCTGAAA 608
    |||
184 rgl1..... 185
609 TCGTGGCAACATCGAAGACATGATTTGGCGGCGCCGATCTGCGGTT 658
    |||
186 .....VALATGPrProAtArgLyValAArgVa 194
659 TGAGTGCAGCGACATTTCA.....TTTCATGACGCGGTGCGCGCG 699
    |||
194 lArdgHhSAlaAgLyPrOgLyPrOleuAlaHhSArg...HhSArgdAga 210
    |||
700 AATTAACCGGTGTGACCATATTCAGATGTAT.....TAC 740
    |||
210 sPlAlArgrArgrAlaAlaHhSArgLeuArgrAlaAspAspAlaGlnVal 226
    |||
741 CATTTGGCGGTTTGT...TGCAACAGCGCGTCTGAACGCGCGCGTGA 787
    |||
227 HhSAlaPrOgLyValArgrHhSArgLeuArgrAlaHhSArgLyHhSgl 243
    |||
788 TTGC.....CTAGGTGTGTTCAGTCAACACAAACGCGCTCTTG 828
    |||
243 yValGlnValGlnGlnArgrAlaValArgrProAspHhSleuAlaAspLeuA 260
    |||
829 CGTACCGTTTGGGTGCGAAGATGCGCAATTTACGCGCGGGAATGTGT 878
    |||
260 rgsAspArgrValGly.....ValAlaGlyArgr..... 268
    |||
879 TGACACAGACACCGCGGTGATTTCCGTTGCGGTATTCAGCGCGCGATTA 928
    |||
269 .....ArgLeuGlyVal...HhSglngLyHh 276
    |||
929 CACAGCGCGCGACGATTTATTTGGACGCTACACAA.....T 966
    |||
276 sglValaAspLeuArgrValrHhSglngLyngLyPrOAlaAspLeuGlyG 293
    |||
967 CAGATTTCCGTTTCAAGAAGCGCGACGAAGACGCTGCGGTGCGT 1016
    |||
293 lAspAspLyArgrValGlnArgrHhS..... 300
    |||
1017 TGGCGCGCGCGACGCAATATCTCAT...CACGCGTACACCGCTGCGGC 1063
    |||
301 .....GlyGlnValValHhSglArgrAlaAlaValArgrProInPr 313
    |||
1064 ATTTCCTGAAGAAACAACCTCTCAAGTTCAACACAGCGCGTCAACGG... 1109
    |||
313 oValAlaGlnGlyLeuPrOValAlaArgSerArgrAspAspValGlnArgrAsp 330
    |||
1110 ..CGCGCGCGCGCGCATGTCGCGATTTGACTTACGCGCGGTGATGCC 1157
    |||
330 lylAlaGlyPrOArgrGlyGlyAlaAspGlyAlaLeuGlnArgrGlnGlnArgr 346
    |||
1158 CTGTGATATCTCTGCCACACCGCTGTTTGGC 1187
    |||
347 lValAlaLeuHhSAspHhSValLeuAla 356

```

seq_name: sp_r0dent:061191

```

seq_documentation block:
ID 061191 PRELIMINARY; PRT: 2045 AA.
AC 061191;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE TRANSCRIPTION FACTOR C1 (HCF).
GN HCFCL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.

```

```

RC TISSUE=LIVER;
RA Kristie T.M., Dasher R.;
RT "CDNAS" encoding the mouse homolog of the human transcription factor C1
  (HCF).";
RL Submitted (APR-1996) to the EMBL/GenBank/DBD databases.
DR EMBL; U53925; AAB01163.1; -.
DR MGI; MGI:105942; Hcfcl.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR001798; Kelch.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF01344; Kelch; 4.
DR SMART; SM00060; FN3; 1.
SQ SEQUENCE 2045 AA; 210537 MW; B89CC2FDA35B969F CRC64;

```

alignment_scores:
 Quality: 109.50
 Ratio: 0.423
 Percent Similarity: 41.977
 Percent Identity: 19.935

alignment block:
 US-09-303-518D-125/rev x 061191

Align seg 1/1 to: 061191 from: 1 to: 2045

```

1343 CAGCCTTCCTCTCATGCTTTCAGACACTTTCGCAACAGCGCGCTA 1294
    |||
747 lysPrOthrHhSleuGlyHhSleuSerVal.....SerProSe 759
    |||
1293 TTTCGTATTTCCCGCGCGACAGAGAGTGCACAAAGCGTCTTCGT 1244
    |||
759 rThrHhSPrOgLyHhSPrOThrHhSleuHhSPrOThrHhSleuHhSPrOThrHhS 776
    |||
1243 CCAATTCAGCAAGCAACCAATGCTGCGCGCTGCGTATCGCCGATTT 1194
    |||
776 lAllelHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhS 792
    |||
1193 AATTCG..... 1188
    |||
793 lYserPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhS 809
    |||
1187 .....CGCAAAAGCAGCGGTGCGCGCA 1166
    |||
809 yAlaPrOAlaHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhS 826
    |||
1165 TATCCAAAGCGCATCAGCGC..... 1146
    |||
826 lYglnGlnGlyValHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhS 842
    |||
1145 .....TCGTAGTACCAATGCGACCATGCGCGCGGTGCGCGCGC 1108
    |||
843 GlyThrHhSleuArgrHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhS 859
    |||
1107 GTTTCAGC.....GCTGTGTGAACCTTGAAAGTTTGT 1076
    |||
859 oValThrValSerAlaValHhSPrOAlaValHhSPrOAlaValHhSPrOAlaValHhSPrOAlaValHhS 872
    |||
1075 TTTCAGAAATAGCGCGAGGTGTGACG..... 1047
    |||
873 .....ValValHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhS 882
    |||
1046 .....GTGANGAGTATTTGCGCGCTGCGCGCGC 1018
    |||
883 lGlyThrValHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhS 899
    |||
1017 AACCAGCGCAAGCGCTTTCGCGCGCTTTCGATTAAC.GGAATTC 969
    |||
899 hHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhS 916
    |||
968 TGATGTGTAGCGTCCCAATATTC..... 943
    |||
916 lAlaThrHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhSPrOThrHhS 932
    |||

```

```

943 ..... 943
933 AlaAlaInThrLeuThrAlaAlaGlyLeuThrProThrII 949
943 ..... 943
949 eThMetGlnProValSerGlnProThrGlnValThrLeuIleThrAlaP 966
942 ..... GTGGCGGCGTTG 931
966 roSerGlyValGluAlaGlnProValHisAspLeuProValSerIleLeu 982
930 TGTAATGCGCGCGTTCAATACGACCC..... 904
983 AlaSerProThrThrGlnGlnProThrAlaThrValThrIleAlaAspSe 999
903 ..... GGAAATACAGCGGGTCTGTCTGCATACC 876
999 rGlyGlnGlyAspValGlnProGlyThrValThrLeuValCysSerAsnP 1016
875 AATTGCGCGCGAGTAATTTGCGATACTTTCGACCCAAACGCTACGCGAA 826
1016 roProCysGluThrHis.....GluThrGlyThrThr 1026
825 GAGCGCGGGTTGTGACTTGAGAACACCTAGGCGAATAC.....GC 782
1027 AsnThrAlaThrThrThrValAlaAlaAsnLeuGlyGlnHisProGlnPr 1043
781 GCTCGGTGTTCAGACGCGCTGTTCGAACAA...ACGGCG..... 745
1043 oThrGlnValGlnPheValCysAspArgGlnGluThrAlaAlaSerLeuV 1060
744 .....AATGTAATTTACATCTGTAATTCAT 718
1060 alThrSerAlaValGlyGlnGlnAsnGlyAsnValAlaValArgValCysSer 1076
717 GGTCCACAGCGTTTATTTCGCGCGCGCGCGCTGTAAGATGTCGCG 668
1077 AsnProPro.....CysGluThrHisGluThrGlyThrAsnThrAl 1091
667 TGCCACTCAACCGCGAGAGATCGG.....GCCGCGCAATTCGA 630
1091 aThrThrAlaThrSerAsnMetAlaGlyGlnHisGlySerSerAsnProp 1108
629 TGTGT.....TTGATGTTGGCAGCATTTTCAACAGCGACGCTGCGCC 586
1108 roCysGluThrHisGluThrGlyThrThr.SerThrAlaThrThrAlaMe 1124
585 AGCTGCTTACAAACATGATTTTGGCTGCGTCAACAGCGCTCAATTCGA 536
1124 tSerSerMetGlyThrGlyGlnGlnAlaGlyAspThrArgArg...ThrThrA 1140
535 ACAGGCGCG.....CGTTGAATCTCGCGCGCTCTTGATGTAATG 495
1140 snThrProThrValValArgIleThrValAlaProGlyAlaLeuGluArg 1156
494 ACCGTAAGGCTGCGCA.....GCCAGCGGATTTGCT 466
1157 ValGlnGlyThrValLysProGlnCysGlnThrGlnGlnThrAsnMetH 1173
465 GTCCATCGCATGACGAGATGGGAGACGCGTGGCATGACGCGAGGAA 416
1173 rThrThrThrMetThrValGlnAlaThrGlyAlaProCysSerAlaGlyP 1190
415 TTTTGCGAAGCGAGCGGTG.....CGCAGCGCATCACAACCGCAT 372
1190 roLeuLeuAlaArgProSerValAlaLeuGlnSerGlySerHisSerProAla 1206
371 TGGATCAGGTTTGGCGGCACTTCTGCGCGCTTAAGTTGCGACGCGCTTC 322
1207 PheValGlnLeuAlaLeuProSerValArgValGlyLeuSerGlyProSe 1223
321 AGGTGGCTAGCGTTCAAACTGATTTGCTGCTTCACCTTCACGGAATCA 272

```

```

1223 rSer.....LysAspMetProThrAlaArgGlnProGluThrThrIleST 1238
271 CGACTGACTGAGTACGCGCGCTTTCCGACAGGTGAATCGCGCGATTTTG 222
1238 hTyThrThrAsnThrProThrThrThrArgSerIle...MetValAla 1253
221 CTTGAAGCGCGCGCAGTAACACACGCGCGATTTCTTTTCTTCGCAAA 172
1254 GlyGluLeuGlyAlaAlaArgValAlaPro.....Th 1264
171 CAGCAGCTTGCGCTTTTTCAGCGCA...TCGCGCTTCCTTACTTCATCG 125
1264 rSerThrTyGlnSerLeuGlnAlaSerSerProSerSerThrMetIrm 1281
124 AGGCGCGCATACCGCATATTTCTGCGCACG..... 93
1281 eThrAlaLeuGluAlaLeuLeuCysProSerAlaThrValThrGlnVal 1297
92 .....AACGC 88
1298 CysSerAsnProProCysGluThrThrHisGluThrGlyThrThrAsnThrAl 1314
87 GACTTGGTAATAGCGCGCGCGCTGTAACAGCGCTTCTCGCGTGTGCGCC 39
1314 aThrThrSerAsnAlaGlySerAlaGlnArgValCysSerAsnProPro 1330

```

seq.name: sp_trodent:Q9QWH2

seq_documentation_block:

```

ID Q9QWH2 PRELIMINARY; PRT; 2045 AA.
AC Q9QWH2.
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE C1 TRANSCRIPTION FACTOR.
GN HCF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Kristie T.M.;
RT "cDNAs encoding the mouse homolog of the human transcription factor C1
(RCF).";
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U08021; AAD09225.1; -.
DR MGD; MGI:105942; Hcf1.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR001798; Kelch.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF01344; Kelch; 4.
DR SMART; SM00060; FN3; 1.
SQ SEQUENCE 2045 AA; 210546 MW; 7AD38FCD78EAB9C CRC64;

```

alignment_scores:

```

Quality: 109.50 Length: 617
Ratio: 0.423 Gaps: 25
Percent Similarity: 41.977 Percent Identity: 19.935

```

alignment_block:

US-09-303-518D-125/rev x Q9QWH2 ..

Align seg 1/1 to: Q9QWH2 from: 1 to: 2045

```

1343 CAGCTTCCCTTCATATGTTTCACACTTTGCGAACAAGCGGCGCTA 1294
:::||||::: |||::|:::
747 LysProThrThrLeuGlyIleSerSerVal.....SerProSe 759
1293 TTGCTATTGTCGCGCGCAAGAGCTGCACAAAGCGTCTTCTTGCT 1244
::: ||||| ||| ::| ||::: |||

```

```

759 rthrhlypsprogllythrthrlllelllelysthrllleprometsera 776
1243 csaattccaaagccaaaccsaatcctgcgcgctgtcgatgcsagacatt 1194
      ::      |||  ::|||  :::::  |||||  |||
776 lallelthrcinalaglylathrclyalthrserserproglylde 792
1193 aatgc..... 1188
      |||||
793 lysserprollethrlllelthrthlysvaalmethlserylthcl 809
1187 .....ccgaaagcagggtgcgacga 1166
      |||  ::|||::
809 yalaproalalylllelthrlnalavalprolyslllealthrlynihg 826
1165 tamccaaagggcatcacggc..... 1146
      :::::|||||:::
826 lylnclnglylvalthrlnvalvalleulysglyalaproglnglnpro 842
1145 .....tcgaatgtaacatcgccacatgcgcggcgccgccc 1108
      |||||::|||  ::  ||
843 glythrllleulnarghrthralpromecllylvalargyeulvalthrrp 859
1107 gttgacg.....gcgtgctgaacttgacagatttgt 1076
      |||||  ::|||
859 ovalthrvalseralavallysproalvalthrrheu..... 872
1075 ttttcagaaatgcgcgaggggttgacg..... 1047
      |||||:::
873 .....valvallysglythrthnglyalthrthr 882
1046 .....gtgatgagatattgttcgcgcgcgcgcgc 1018
      |||  |||||::|||::|||
883 leu glythrvalthr glythrvalserthrserleuhalaglyal 899
1017 aaccacccgaacagcctctgcgcgccttctgcataac ggaatc 969
      |||  :::::|||||  |||  :::::
899 ahlsertthrserlserleualathrprolethrleu glythr 916
968 tgatttgtagcgctcccaaatatc..... 943
      |||  :::::|||||  |||  :::::
916 lealathrleuserserglnvalleasnprothralllethrvalser 932
943 ..... 943
943 alalaglnthrthrleuthralalaglyleuthrthrprothrll 949
949 ethrmetslprovalserglnprothrlnalthrleullethrlnar 966
942 .....gtgcggcgccctg 931
966 roserglyvalglualaglnprovalniaspleuprovalserlleu 982
930 tgatatccgcgcgcttcgaatccgacac..... 904
      :::::|||||  |||
983 alaserprothrthnglnprothrlnalthrleualaspse 999
903 .....ggaatcacgcgctgtctgttcacac 876
999 rcllynglyaspyalaglnprogllythrvalthrleuvalcysserasnr 1016
875 aattgcgcgcgagaaatttcgcatacttcgcacccaaacacgtagcga 826
1016 roprocgsluthnhs.....gluthr glythr 1026
      :::::|||||
825 gagggcggttgtagcttcagacaccactagggcaatcac.....gc 782
1027 asnthralatthrthrthvalvalalasnleuglylniasproglnr 1043
781 gcctgcggttgtagcagcgctgttcacaaacaa...acggc..... 745
      |||||  ::|||  ::|||
1043 othrlnalaglnrhevalcysaspragnglgnluthralalaserleu 1066

```

```

744 .....AATGGTAATTACACTTTGAAATTTGAT 718
1060 aAlhrSeAlaValaIcIyngIngnIngnIaYsnValaIaArValCysSer 1076
717 GGGCCACAGCGGTTTATTCGGCGCCAGCGGCTCGATGAATGAATGTGCG 668
1077 AasnProPro.....CysgluThrHnIsgluThrIcIyThrHnAsnThAl 1091
667 TGGCATTCAAACCGGACGAGATGGCG.....GGCGCGGAAATTA 630
1091 aThrThAlaIaThrSerAsnMetAlaIcIyngInnIsIySerSerAsnPro 1108
629 TGGGT.....TTGCATTTGGCAGCAATTTTCACAGCGCACGTCTGGCC 586
1108 roCysgluThrHnIsIcIyngInhrGlyThrThr..SerThAlaThrHnAla 1124
585 AGCTGCGCTTACAAACATGAGATTTTCGCTGCGTGAACGGCTCAATACAA 536
1124 tSerSerMetglYhrngIcIyngIngnIaArAspThArVala...ThThAla 1140
535 ACAGGGCG.....CGTTGAATTCCTGGCGGCGCTTCTTGATTAATG 495
1140 snThrProThrValaIaIaArgIeThrValaIaProglYalAlaIeGluArg 1156
494 ACCGTAAAGGTCGGCA.....GCCAGGGATGGT 466
1157 ValGIngnIyThrValaIySProInCysIcIyngIngnIcIyngInnAsnMetTh 1173
465 GTCCATGCGCATTCAGCAAGATGGCGACAGCGCTCGGCATCGACGGCAGAA 416
1173 rThrThrThrMetThrValGInnIaThrGlyAlaArroCysSerAlaIyP 1190
415 TTTTGGTGAACGAGCGGGTg.....CCGACGGCAGTCCAAACCGGAT 372
1190 roIcIyThrAlaArProSerValaIaIaIeGluSerGlySerHnSerProAla 1206
371 TGGATCGAGTGGCGGCGGCACTTCTTCGCGCGCTTAAGTTGGCAAGCGCTTC 322
1207 pheAlaIcIyngInnIaIeUrProSerValaIaArValaIcIyIeUserGlyProSe 1223
321 AGGTGCGTACCGTTCAAACTCCATTTTCGTCGGCTTCACAGCGCAATCA 272
1223 rSer.....LysAspMetProThrAlaArgInProgluThrTyHnIst 1238
271 CGACTGCAATGATACGCGGCTTTTCCGACGGTGAATCGCGGCGAATTTTG 222
1238 hrTyHrThrHnAsnThrProThrThrThrArgSerIle...MetValaIa 1253
221 CCTGAGCCGCGCGCAGTAAACACACAGCGCGGAGATCTTTTGTCTTCAA 172
1254 GlyIcIyngInnIaIaIaArValaIaValPro.....Th 1264
171 CAGACATTTGGCGCTTTTGTGAGCGCA...TGCGCTTCCTTGACTTTCATCG 125
1264 rSerThrTyIcIyngInnIaIeUrProSerValaIaSerSerProSerThrMetThrm 1281
124 AGGGGCGCATCGCGCATATTCTTCCGCACAG..... 93
1281 etThrAlaIeGluAlaIaIeUcysProSerAlaThrValaThrGlnVal 1297
92 .....AACGC 88
1298 CysSerAsnProProCysgluThrHnIsIcIyngInhrGlyThrThrAsnThAl 1314
87 GACTTCGGTAATGCGCGCGCGCGCGTGAACAGCGCTTCCTCGGCTGGCC 39
1314 aThrThrSerAsnAlaGlySerAlaGlnArValaCysSerAsnProPro 1330
Seq_name: sp_rVirus_Q98MW2
Seq_documentation_block:
ID Q98MW2 PRELIMINARY: PRT: 1504 AA.

```


AC 098W2;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE GAG POLYPROTEIN [CONTAINS: CORE PROTEIN(S) P24].
 GN GAG.
 OS Avian leukosis virus.
 OC Viruses; Retrovirdae; Retroviridae; Alpharetrovirus.
 ON NCBI_TaxID=11864;
 RX MEDLINE=21165271; PubMed=11264350;
 RA Johnson J.A., Heneline W.;
 RT "Characterization of endogenous avian leukosis viruses in chicken
 RT embryonic fibroblast substrates used in production of measles and
 RT mumps vaccines."
 RL J. Virol. 75:3605-3612(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ENDOGENOUS AVIAN LEUKOSIS VIRUS 1, EV-1;
 RA Johnson J.A., Heneline W.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO GAG GENE PROTEIN P24 (CORE NUCLEOCAPSID PROTEIN).
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO
 CC KNOWN AS THE RETROPEPSIN FAMILY.
 DR EMBL: AY013303; AAK13200.1;
 DR HSSP: P03322; 2RSP.
 DR InterPro: IPR001969; Asp.-protease.
 DR InterPro: IPR001995; Asp.-protease.
 DR InterPro: IPR000721; Gag-P24.
 DR InterPro: IPR003308; Integrase_Zn.
 DR InterPro: IPR004028; Retro_M.
 DR InterPro: IPR002156; RNaseH.
 DR InterPro: IPR001584; Rve.
 DR InterPro: IPR000477; RVTse.
 DR InterPro: IPR001878; Znf.CCHC.
 DR Pfam: PF00607; gag_P24_2.
 DR Pfam: PF02022; Integrase_Zn; 1.
 DR Pfam: PF02813; Retro_M; 1.
 DR Pfam: PF00075; RNaseH; 1.
 DR Pfam: PF00665; Rve; 1.
 DR Pfam: PF00077; RVP; 1.
 DR Pfam: PF00078; rvt; 1.
 DR Pfam: PF00098; zf-CCHC; 2.
 DR SMART: SM00343; Znf.C2HC; 2.
 DR PROSITE: PS00141; ASP_PROTEASE; 1.
 DR PROSITE: PS01175; ASP_PROT_RETROV; 1.
 KW Aspartyl protease; Core protein; Hydrolyase; Polypeptide;
 KW RNA-directed DNA polymerase; Zinc-finger.
 FT CHAIN 1 175
 FT CHAIN 178 239
 FT CHAIN 240 479
 FT CHAIN 489 577
 FT CHAIN 578 701
 FT CHAIN 709 1504
 FT CHAIN 709 1280
 FT CHAIN 1281 1504
 FT CHAIN 1504 AA; 163199 MM; 979779BA89620435 CRC64;
 SQ SEQUENCE
 alignment_scores:
 Quality: 107.50 Length: 486
 Ratio: 0.500 Gaps: 32
 Percent Similarity: 44.239 Percent Identity: 22.634
 alignment block:
 US-09-303-518d-125 x 098W2 ..
 Align seg 1/1 to: 098W2 from: 1 to: 1504
 9 AATCAAAAAGGTCTAAGCTGCGCATCGCGGAGACCGGAGCAAGCCG 58

492 AsnArgLuarGasp.....GlyGlnThrGlySerIyl 503
 59 TTTACGAGCG.....CCCGGCATATACGAA 84
 503 YArgAlaArgGlyLeuCySTyTrhCysGlySerProGlyHISLyGlnA 520
 85 GTCCGTTGCTGGCGAAGATATGCGGATGCGCCCTCGATGAAGT 134
 520 La..... 520
 135 CAAGAAAGCGATGCGCTCAAAAAGCCAGTGTGTTGAACAA 184
 521GlnCysProLysArgLysSerGlyAsnSerArgLyu.. 533
 185 AGATCCGGCGGTGTACTGCGCCGCTTCAGGCAAAATCCCGCG 234
 534ArgCysGlnLeuCySaspGlyMetGlyHISAsn..... 544
 235 ATTACCGTGGCGAAGCGCGTACTTCACTGATGATGCGCGTTGA 284
 545AlaLysGlnCysArg... 549
 285 AGCAACGAGCAAGATGAGTTGAACGCTACGACCTGAAGCGCTGCA 334
 550ArgArgasp.....Gly.A 554
 335 ACTTAAGCGCGAAGAGTGGCGCGCAACCTGATCCATCCGGTTGCG 384
 554 sn...GlnGlyGlnArgProGlyArgGlyLeu...SerSerGlySerTrp 568
 385 ACTGCGCTGCGACCGCTGCTTCAGCAAAATCTCCGCTGCAATGCGCA 434
 569ProValSerGlnGlnProAlaValSer..... 577
 435 GCCGTTGCGCATCTGTCATGCGATGCAATCCGCTGCTGCCG 484
 578 ...LeuAlaMetTrhMetGlnHISLysasp...ArgProLeuValArg. 591
 485 ACCCTAGCGTCATTATCAAGAGAGCGCGAGATTTCAAAGCGGCGCTG 534
 592ValIleLeuThrAsnThrGlySerHisProValLysGlnArgSer 606
 535 TTGGATTGAGCGCTTTGACCGAAGCAAAATCCATGTTGAAGCAGC 584
 607 ValTyrIleThrAlaLeuLeuasp.....Se 615
 585 TGGCGACAGCGT.....C 598
 615 rGlyAlaAspIleThrIleLeuSerGlnGluAspTrpProThrAspTrp 632
 599 CGTGTGAATATGTCGCAACATCGAATGATGCGCGCGCGCAT 648
 632 roValMetGlnAlaAlaAsnProGlnIleHisGlyIleGlyGlyIle 648
 649 CTGCGCGTTGAGTGGCGACGACATTCATTTCATGAGCGGTCGCGCG 698
 649 PromethrGlySer..... 653
 699 GAATAAACCGTGTGACATCATTAATCAAGATGTAATACCATTTGGC 748
 654ArgAspMetIleGluValGly 661
 749 GTTGTGTTGCAACAGCGCTGCAACAGCGCGTGTATCCCTAGGT 798
 661 alIleAsnArgaspGlySerLeuGluArgProLeuLeuLeuPheProAla 677
 799 GGTTCACAGTCAACAA.....CCGCGCTTGGTGGTGGTGG 842
 678 ValAlaMetValArgArgGluTyrProArg..... 687
 843 TGGGAAGTATGCGCAATTTACGCGGCGAATGTTGTTGACACAGCAAC 892

```

688 .....LysArgLeuSerAlaGlyProArgAlaProLeuAspLysP 701
893 GCGTATTTCCGGTTCGATTTGAAGCGCGGATACAAAGCGGCAC 942
701 heilegilyargAlaThrValLeu..ThrValAlaLeuHisLeuAlaIlePr 717
943 GATTATTTGG.....GACGCTACCAATC.....AGATTTCGT 977
717 oleuylstrLysProAspHisThrProValThrIleAspIleThrProL 734
978 TATGAGAAAGCGCC.....AGCAAGAGC 1003
734 eu..ProGluGlyLysLeuValAlaLeuThrGluLeuValGlyLysGlu 750
1004 TG...TTGCGTGGGTGGCGCGCAAGCGCAAAATCTCC..... 1041
750 euGluLeuGlyHisIleGluProSerLeuSerCysThrAsnThrProVal 766
1042 ...ATCAGCGGTACAAACCTCGCCATTCTGAAAAACAACTCTTCAA 1088
767 PheValIleArgLysAlaSerGlySerTy.....ArgLeuLeuH 780
1089 GTTCAACACAGCGCGTCAAGCGCGCGACCGCCATGGTCCGATGGTA 1138
780 SASPLEuAlaGAlaValAlaAsnAla.....LysLeuValIleProPheGly 794
1139 CTACAGAGCGCGTATGCC...TTGATATCTCTGCC.....ACC 1176
794 lAlaValGluGlnGlyAlaProValLeuSerAlaLeuProArgGlyTrpPro 810
1177 CTGCTTTGGCGGATTTAATCTCGCGGATACGACGCGCGCATG 1226
811 LeuLeuValLeuAspLeu.....LysAspCysPh 820
1227 GCGTTCGTGGAATTGGACGAGACAGACTCTTGTGACGCTTCTGT 1276
820 ePheSerIleProLeuAlaGluGlnAspArgGluAlaPheAlaPheThrL 837
1277 GCCCGGCAATACGATACGCGCGCGCTGTGGC.....AAGTG 1317
837 euProSerValAsnAsnGlnAlaProAlaArgArgPheGlnTrpLysVal 853
1318 CTG 1320
854 Leu 854
seq_name: sp_bacteriap:Q97PQ2
seq_documentation_block:
ID 097PQ2 PRELIMINARY; PRT; 914 AA.
AC 097PQ2:
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CATION-TRANSPORTING ARPASE, EI-E2 FAMILY.
GN SPI51.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
MEDLINE=21357209; PubMed=11463916;
RA Tettein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouli H., Wolf A.M., Uteback T.R., Hansen C.L.,
RA McDonald L.A., Feldlyum T.V., Anguoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RA *Complete genome sequence of a virulent isolate of Streptococcus

```

```

RT Pneumoniae".
RL Science 293:498-506(2001).
DR EMBL; AE007449; AAK75638.1; -.
DR TIGR; SPI51; -.
DR InterPro; IPR004014; Cation_ATPase.
DR InterPro; IPR001757; EI-E2_ATPase.
DR InterPro; IPR000695; HATPase.
DR InterPro; IPR001454; Hydrolyase.
DR InterPro; IPR000661; Na_H_K_ATPase.
DR Pfam; PF00689; Cation_ATPase_C; 1.
DR Pfam; PF00690; Cation_ATPase_N; 1.
DR Pfam; PF00122; EI-E2_ATPase; 1.
DR Pfam; PF00702; Hydrolyase; 1.
DR PRINTS; PR00119; CATATPASE.
DR PRINTS; PR00120; HATPASE.
DR PRINTS; PR00154; ATPASE_EI_E2; UNKNOWN_1.
DR PROSITE; PS00154; ATPASE_EI_E2; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 914 AA; 99051 MW; C86323039FE51850 CRC64;

```

```

alignment_scores:
Quality: 107.00 Length: 356
Ratio: 0.615 Gaps: 15
Percent Similarity: 48.876 Percent Identity: 23.315

```

```

alignment_block:
US-09-303-518D-125 x Q97PQ2 ..

```

```

Align seg 1/1 to: Q97PQ2 from: 1 to: 914

```

```

28 CTGCCCCATCGCGGCGAGACCGGACGAGCCGTTTACGAGCGCGCCCAT 77
111 ::::|||||:||||| |
303 LeuAlaValAlaIleProGluGlyLeu.....ProAlaIle 315
78 TACCGAAGTGCCTTG...CTTGCGAGAGATATGCGCGTATGCGCCCT 124
111 ::::|||||:||||| :
315 eValThrIleValLeuSerLeuGlyThrGlnValLeuAlaLysArgHis 332
125 CGATG.....AAGTCAGAGAGCGGATGCGCTCAAAAAAGCGCAAGG 168
111 ::::|||||:||||| :
332 erIleValArgLysLeuProAlaValAlaGluThrLeuLysSerThrGluIle 348
169 CTGTTGAGACAAAGAAAGANT.....CC 191
111 ::::|||||:||||| :
349 IleAlaSerAspLysThrGlyThrLeuThrMetAsnLysMetThrValG 365
192 GCGCGTGTGTTTACTGCGCGGCTTCAGGCAAAATCGCGCGATTCACC 241
111 ::::|||||:||||| :
365 uLysValPheTyArgAlaValLeuHisAspSerAlaAspArgIleGlu 382
242 GTGGC...GAAAGCGGCTACTTCAGTCAGTCGATGCGCGTGAAGGC 288
111 ::::|||||:||||| :
382 euGlyLeuGluMetProLeuLeuArgSerValValLeuAlaAsnAspThr 398
289 AACGACGAATCGAGTTTGAACGCTACGACCTGAAGCGCTGCAAACTT 338
111 ::::|||||:||||| :
399 LysIleAspValGlu.....GlyAsnLe 406
339 AAGCGCGAGAGAGTGGCGCGCAACCTGATCCAAATCGGTTTGGAGCTG 388
111 ::::|||||:||||| :
406 uIleGlyAspProThrGluThrAlaPheIleGlnTrpAlaLeuAspLysG 423
389 CGCTCGGACCGGCGCGCTTC...AGCAAAATCTCTCGCGTGCATGCGGAG 435
111 ::::|||||:||||| :
423 lYTyArgValLysGlyPheLeuGluLysTyArgValGAlaGluLeu 439
436 CGGTTCGCGCATCTTCGTCATGCGATGAGACAC...AATCGCTGGGTGC 482
111 ::::|||||:||||| :
440 ProPheAspSerAspArgLysLeuMetSerThrValHisProLeuProAs 456
483 CGACCTTACGTCATTAACAAGAGCGCGGAGATTTCAAGCGCGCC 532
111 ::::|||||:||||| :

```

```

456 pGlyArgPheLeuValAlaValLysGlyAlaProasp..... 468
533 TGTGTATTGAGCGGTTTGACGAGCAAAATCCATGTTTAAAGCA 582
469 .....GlnLeuLeuLysArgCysLeuLeuArgAspLysAla 480
583 GCTGGCCAGACGTCGCGCTGAAATGCGCCAC...ATGGAACACA 629
481 GlyAspIleAlaProIleAspGlyLysValThrAsnIleIleArgThrAs 497
630 TGAATTGGGGGGCCGATCTGCC.....GGTTGAGTGGCAGGCACA 673
497 nasSerGlnMetAlaIleGlnIleAlaLeuArgValIleAlaGlyAlaTyr. 513
674 TTCAATTTCATGAGCCGTCGCGCGCATTAACCGTGGACCCAT 723
514 ..LysIleIleAspSerIleProGluAsnLeuThrSerGluLeuGlu 529
724 TATCAAGATGTAAATTACCATGGCCGTTGTTGCACAGCCGCTGAA 773
530 ...AsnAspLeuIlePheThrGlyLeuIle.....GlyMetIleAs 542
774 CACCGAGCCGTCGATGACCCCTAGTGGTTCGACATCAACAACGCCGCC 823
542 pProGluArgProGluAlaIleAlaGlyAlaValArgValAlaLysGluAlaG 559
824 TCTTGGCTACCGTTTGGGTGGCGAAAGTATCGCAATTAATCTGGCGGCA 873
559 LylLeuArgProIleMetIleThrGlyAspHisGlnAspThrAlaGluAla 575
874 TTGGTT.....GACACAGCAACGCCGCT 896
576 IleAlaLysArgLeuGlyIleIleAspAlaAsnAspThrGluGlnHisVa 592
897 GATTTCGCGTTCGATTGAACGCGCGCATTAACAAGCGCGCAGCAT 946
592 IleuThrGlyAlaGluLeuAsnGluLeuSerAspGluIlePheGluLysV 609
947 ATTGGAGCGTACCAACAATCAGATTTCCGTTATGAGAAGGCCGAGC 996
609 alValGlyIleIleThrSerValTyrAlaArgValSerProGluHisLysVal 625
997 AAGAGCTGTCGCGCTG 1014
626 ArgIleValLysAlaIlePro 631

seq_name: sp_human:095769
seq_documentation_block:
ID 095769 PRELIMINARY; PRT; 915 AA.
AC 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE WIGSC:H_DJ412A9.3 PROTEIN.
GN WIGSC:H_DJ412A9.3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RA MEDLINE=9063792; PubMed=9847074;
RA Sulston J.E., Waterston R.;
RT "Toward a complete human genome sequence.";
RL genome Res. 8:1097-1108(1998).
RN 12
RP SEQUENCE FROM N.A.
RA Murray J., Lennon S., Harmon G.;
RT "The sequence of Homo sapiens PAC clone RP3-412A9.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN 13
RP SEQUENCE FROM N.A.

```

```

RA Waterston R.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005005; AAD15619.1; -.
DR HSSP; 001082; 1BRK.
DR InterPro; IPR001715; Calponin_hom.
DR Pfam; PF00307; CH; 1.
DR SMART; SM00033; CH; 1.
DR PROSITE; PS50021; CH; 1.
SQ SEQUENCE 915 AA; 98919 MW; FFOAC36BDEC9B30 CRC64;

```

```

alignment_scores:
  Quality: 107.00      Length: 500
  Ratio: 0.446        Gaps: 31
  Percent Similarity: 48.000  Percent Identity: 24.200

```

alignment_block:

US-09-303-518d-125 x 095769 ..

Align seg 1/1 to: 095769 from: 1 to: 915

```

26 ACCTGCCATCGCGGAGAGCCGAGCAAGCCGTTACGAGCCGCCGCC 75
111 CG.....GATGCGCCGCCGCA 127
231 roGlySerProGluProProSerProProLysThrThrSerProGlu 247
128 TGAAGTCAAGAGAGCGATCCGCTCAAAAAGGCGCAAGTGTGTTGAA 177
248 ProGlnGluSerProThrLeuProSerThrGluGlnValValAsnLys 264
178 G..... 178
264 sLeuLeuSerGlyProLysGluThrProAlaAlaGlnSerProThrArg 281
179 .....ACAAAAGANTC..... 190
281 LProSerAspThrLysArgAlaAspValAlaGlyProAlaGProCysGln 297
191 CGGCGTGTGTTACTG.....CGCGGCTTCAGCAAAATCGCGCG 234
298 ArgSerLeuSerValLeuSerProArgGlnProAlaGlnAsnArgGln 314
235 ATTCAACGTCGCGGAAAGCGGACTTCAGTCAAGTGTGTTCCGTTGA 284
314 erThrProLeuAlaSerGlyProSerSerPheGlnArgAlaGlySerVal 330
285 AGGCAACAGCAAGAAATCGATTGAGCTAGCGACCTGAGCGCGGCA 334
331 ArgAspArgValHisLys..PheThrSerAspSerPrometAlaAlaArg. 346
335 ACTTAACGCGGCAAGAAAGTGGCGGCAACCTGATCCATCCGCTTG... 381
347 .....LeuGlnAspGlyThrPro 352
382 TGAACGTGCGTGGCAACCGTTCGTTACGAAATTCCTG...CCGTGA 428
353 GlAlaAlaLeuSerProLeuThrProAlaArgLeuGlnGlyProSerL 369
429 TGCCGAGCGGTCGTCATCTGTGATGCGATGACACCAATCGCTGG 478
369 euthrSerThrThrProAlaSerSerSerSerGlySerSerArgG 385
479 C...TCCCGACCTACGTCATTAACAAGAGCGCGAGGATTTCAA 525
385 yProSerAspThrSerSerArgPheSerLysGluIleArgGlyValAlaG 402

```

```

526 CGGGCCCTGGTATGACCGGTTGACCGACGCAAAATCATGTTG 575
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
402 InProleuAlaInleuArGserGProGInGluGluPro.ArGgl 418
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
576 TAAGCAGCTGGCGGACGAGCTGGCTGAAAATGCTCCAC... 618
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
418 yArgGlyLeuAlaAlaArg...ProleuAlaInArGAlaGlyLeuPro 434
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
619 ..ATCGAAACATGAAATCGGCGCGCATCTGCGGTTGAGTGGC 666
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
434 alAlaArGserGluGluProGAlaProleuProValAlaVal...Gly 449
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
667 ACCGACATTCATTCATCGACCGGCGCGGCAATTAACCGGTGAGC 716
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
450 Thr.....AlaGluProGlySerMetLysThrThrPheTh 462
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
717 CATCATTTATCAAGATGTAATTAACATTTGGCGTTGTTGACAGCGC 766
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
462 rllleGluLleLysAsp.....GlyArg.....GlyG 471
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
767 GTCGAAACACGAGCGCGTATGCGCGTGGTGTTCACATCAACAA 816
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
471 lmlaserThrGlyArgValleuLeuProThrGlyAmGlnArGAlaGlu 487
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
817 .....CCGCGCCCTCTGCGGTACCGGTTTGGG 842
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
488 leuThrleuGlyLeuArGAlaProProThrleuLeuSerThrSerSerl 504
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
843 TGGCAAGATTCGCAATTTACTGCGGCGAATTTGTTGACAGACAAAC 892
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
504 yGlyLys...SerThrleuThrArGAlaInSerProGlyThrleuAla 520
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
893 GCGTATTCGCGTTCGATTCAGACGCGCGATTACACAGCGCGCAC 942
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
520 rgleu.....GlySerValThrHis.....ValThrSerPheSerHis 532
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
943 .....GATTTATTTGGAGCGCTACCAATCAGATTT.....C 974
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
533 AlArProSerSerArGly.GlyCysSerLleLysMetGluAlaGluP 549
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
975 CGTTATCGAAGAGCGCGACGCAAGAGCTGTGGCTGGGTTGCGCGC 1024
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
549 roAlaGluProleuAlaAlaValGluAlaAlaInGlyAlaGluIn 565
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1025 AGCGGACAAAT...ACTCATCGCGGTACAAACCGCGCATTTCCG 1071
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
566 ThrArGValAlaInLysAlaPro.GluGlyArGserProleuSerAlaGlu 582
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1072 AAAAAACAATCTTCAA..... 1088
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
582 luleuMetThrleGluAspGluGlyValleuAspLysMetLeuAspGln 598
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1089 .....GTTCACACAG 1099
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
599 SerThrAspPheGluGluArGlyLysLeu]leArGAlaAlaLeuArGlu 615
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1100 CCGTCAACGGGCGCGACCGCATGTCGCGATTTGTTACTTACGAGCG 1149
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
615 uArGAlaArGlyArGAspGlnArGAspLysGluArGAlaArGLeuG 632
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1150 GTGATCGCTGGATATCCGCCACCGCTGCT.....TTTGGCGCATTT 1193
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
632 lInGluAlaArGlyArGProGlyGluGlyArGAlaInThrAlaThrGlu 648
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1194 AATCGTCGCGATACGACGAGCGCGAGCGCATTTGGG 1229
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
649 ThrThrThrArGHisSerGlnArGAlaAlaAspLys 660
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
seq_name: sp_invertebrate:076602
seq_documentation_block:
ID 076602 PRELIMINARY; PRT; 1275 AA.
AC 076602;

```

```

DT 01-NOV-1998 (TREMUREL. 08, Created)
DT 01-NOV-1998 (TREMUREL. 08, Last sequence update)
DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)
DE HYPOTHETICAL 122.9 KDA PROTEIN.
GN H02F09.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Anscough R., Anderson K., Baynes C., Berts M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
RA Gardner A., Green F., Hawkins T., Laister N., Latille P.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Laister P.,
RA Lightning J., Lloyd C., McMuray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Koopra A., Saunders D., Showken R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Geisel C., Harmon G.;
RT "The sequence of C. elegans cosmid H02F09."
RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF077538; AAC64622.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1275 AA; 122924 MW; CB59B70C05959E25 CRC64;

alignment_scores:
Quality: 107.00 Length: 431
Ratio: 0.500 Gaps: 16
Percent Similarity: 49.652 Percent Identity: 21.578

alignment_block:
US-09-303-518D-125/rev x 076602 ..
Align seg 1/1 to: 076602 from: 1 to: 1275
1319 AGCATTGTGGCAGACAGCGGCGATTCGATTTCCCGGCGACAGCA 1270
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
218 AsnAsnIleuAsnIleGlyThrThrGlnThrProProValThrThrSe 234
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1269 GCTGCACAAAGCGAGCTCTTCGTCCATTCACAGCAACCAAT...G 1223
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
234 rThrMetAlaThrThrThrAlaAsnValThrSerAlaAlaProAsThrT 251
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1222 CCTGCGCGCTGCGATTCGCGCAGATTAATCGCGCAAAAGCAGGTG 1173
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
251 hrValThrIleSerThrSerProThrThrValThrValProSerThr 267
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1172 GCGAGATATCCAAAGGATCAGCGCGCTGTAAGTACCAATCGGACCAT 1123
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
268 AlaGlnThrSerSerThrThrThrValThr...ValProThrThrThra 283
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1122 GCGCGGTCGCGCGCGGTG.....ACGGCTGTGTGAACTGA 1085
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
283 lThrGlyProThrThrValValThrValProThrThrValThrValThr]e 300
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1084 AGAGTTTGTTCAGAAATGCGCGAGGCTGTACCGGTGATGAGATAT 1035

```

```

300 roSerThrValValThrSerProIleThrThrProSerThrValValThr 316
      |||
1034 TTGTCGGGCTTCGGCGACACCGCAACAGCTTTGCTGGCGCTTC 985
      |||
317 ValProSerThrValValThrValProSerThrAlaValThrLysProse 333
      |||
984 TTCGATACGGAATCTGATGTGGTAGCGCTCCCAATATCGTGGCGC 935*
      |||
333 rThrValVal.....ThAlap 339
      |||
934 CTGTGTAAATCGCGCTCAATACCGAACCGGAATCAGCGGTGTCT 885
      |||
339 roSerThrValValThrValProSerThrValValThrLysProSnrThr 355
      |||
884 GTGTACACCAATCGCGCGGAGTAATTCGATACCTTTCGCCACCAAAAC 835
      |||
356 ValValThrSerSerProThrVal.....AlaThrThrProThrTh 369
      |||
834 GTACACGACAGCGCGGCTTGTGACTTGAGAACCCCTAGGCAATCA 785
      |||
369 rValValThrThrProSerThrValValThrValProSerThrValVal 386
      |||
784 CGCGCTCGGTTCAGACGGCGCTTGCAACAAACGGCAATGTAAAT 735
      |||
386 hr.....ValProThrThrValValThrAsnProSerThrValVal 399
      |||
734 ACATCTTGAATATGATGTGACGACATTTTCAGACGCGACGTCGCGCA 585
      |||
400 ThrAla.....ProSerThrValValThrValProThrThVa 412
      |||
684 GATGAATGATATGTGCGTCCCACTCAAAACCGGAGATGCGCGCGCA 635
      |||
412 lMet.....T 414
      |||
634 ATTCAATGTTTCGATGTGCGACGATTTTCAGACGCGACGTCGCGCA 585
      |||
414 hrSerArgSerThrValIleThrThrProThrThrIleGlySerSerPro 430
      |||
584 GGTGCTTCAACATGATGTGCGTTCGCTGCAAAACGGCTCAATACCAA 535
      |||
431 SerThrAlaGlyThr..... 435
      |||
534 CAGCGCGGCTTGAATCTCGCGGCTTTCGATATGACCTAGCGT 485
      |||
436 .....SerLeuAlaSerThrAlaValThrThrGluThrSerIleGlyS 450
      |||
484 CGGACGCGACGGGATG.....GTTCATATCGCATTTAGCAAG 447
      |||
450 eSerSerThrProLeuProSerGlnSerThrSerLeuSerMetSerSer 466
      |||
446 ATGCGG.....AAGGCTCGCATTCAGACGCGAGAGATTTGCTGAACG 403
      |||
467 LeuSerThrThrProSerSerSerThrAlaIleValThrSerProAl 483
      |||
402 ACGGCTGCGACGCGACGATCCAAACAGGATGATGAGTTCGCGGCGCA 333
      |||
483 aThrGlnIleSerThr.....LysProThr..IleGlyThrSerMetS 497
      |||
352 CTCTGCGCGCTTAAGTTGCCAGCGCTTCAGGTGCGTCAAC 303
      |||
497 eSerGlyProThrThrValAlaProIleAlaSerThrGlnSerThrVal 513
      |||
302 TCGATTTCGTGCTTCCTCA...ACGGCAATCAGCATGACGAGTAC 256
      |||
514 LeuGlnSerSerThrProSerGlyThrThrValThrLeuProSerLysE 530
      |||
255 GCGGTTTTCGCCACGCGGATTCGCGGATTTTGCCTGAACGCGGCGAG 206
      |||
530 rSerThrAla.....ThrAlaGlyThrSerProGlnAlaSerThrV 544
      |||
205 TAAACACACGCGCGGATTCCTTTCTTCAACAGCACTTCGCGCTTTT 156
      |||

```

```

544 alThrThrValThrAspIleSerThrValSerLysSerThrValThrSer 560
      |||
155 TTGACGCGATCGCTTCCTTACCTTCATCAGAGGCGCGATACCGGCA 106
      |||
561 GlnThrAlaGlnSerSerLeuSer...ThrGlnSerProThrSerAlaGl 576
      |||
105 TTCCTCCCAAGCAACGCGACTCGGTAATGCGGCGCGCGCTG 63
      |||
576 ySerSerIleSerThrValSerThrValSerGlnProSer 590
      |||

```

seq_name: sp.bacteria:005089

seq_documentation_block:

```

ID 005089 PRELIMINARY; PRT: 545 AA.
AC 005089;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHEICAL 57.3 KDA PROTEIN.
OS Nocardioides simplex (Arthrobacter simplex).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Propionibacteriaceae; Nocardioidaceae;
OX NCBI_TaxID=2045;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO12069;
RX MEDLINE=95319331; PubMed=7596291;
RA Molnar I., Choi K., Yamashita M., Murooka Y.;
RT "Molecular cloning, expression in Streptomyces lividans, and analysis
RT of a gene cluster from Arthrobacter simplex encoding 3-
RT ketosteroid-DELTA-1-dehydrogenase, 3-ketosteroid-DELTA-5-isomerase
RT and a hypothetical regulatory protein."
RL Mol. Microbiol. 15:895-905(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO12069;
RA Dziadek J., Yamashita M., Murooka Y.;
RT "Cloning, sequencing and characterization of the downstream region of
RT ksdBI operon of Arthrobacter simplex."
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z93338; CAB07541.1; -.
KW Hypothetical protein.
SQ
SEQUENCE 545 AA; 57318 MW; 55F441EBD5E3A19E CRC64;

```

alignment_scores:

```

Quality: 109.50 Length: 493
Ratio: 0.464 Gaps: 33
Percent Similarity: 47.870 Percent Identity: 25.152

```

alignment_block:

US-09-303-518D-125 x 005089 ..

Align seg 1/1 to: 005089 from: 1 to: 545

```

18 AGGTCTAAACCTGCCAT...CGGGGACAGCGGAGCAAGCGCTTAGG 64
      |||
61 ArgSerArgGlyAlaGlnValArgGlySerAlaSerArgSerIleYargAr 77
      |||
65 ACGGCGCGGCATTTACGA.....AGTCCGTTGCTTGC 99
      |||
77 gSerSerGlyArgArgArgGlyScyScyProGlyGlySerArgArgArgGly 94
      |||
100 GAAGAATATCGCGTATGCGCCCTCGATGAAGTAAGTAAGAGG...CGA 146
      |||
94 rArgArgArgIleArgValGluPro.....AlaGlnIleYargValasn 107
      |||
147 TGCCTCAAAAAAGGCAAGTGTCTTTGAAGCAAAAAAGATCCGGCG 196
      |||
108 AlaSerSerGlyArgProAlaAla.....ArgAlaSerSerAlaIle 121
      |||
197 TGGTGTCTTACTGC...GCGGCTTACGCAAAATCGCGCGATTCACCGT 243
      |||

```

```

121 g.....CysArgAlaGlyAlaArgSerProArgThrProG. 135
244 GG.....CGAAGCGCGT..... 257
135 lylGlyProSerAlaGlyArgAlaSerArgAlaArgSerCys 151
258 ...ACTTCAGTCAGTCGATGCGTTGAAGCGACGACGAAATCGAGT 304
152 GlyThrThValGlnAspGln.....ProHisArgAlaAspArgVal 165
305 TTGACGCTACGACCTGACGCGTGGCAACTTGAAGCGGAGAAATG 354
165 l.....AlaGlyProValAlaAlaGlnHisGlyL 175
355 CGCGCGACACCTGATCCATC...CGGTTGTGACTGCGCTGCGACCGC 401
175 euProGlnGlyHisValValGlyArgLeuValAspGluProAlaVal 191
402 TCGGTTGACGAAATTCCTGC...CGTCGATCCGAGCGCGTGGCATCT 448
192 AlValAspHisAspArgAlaGlyGlnArgAlaLeuGlyGlnHisLe 208
449 TCGTCATGCGATGACACCAATCCGCTGCGCGACCTGACGCGCAT 498
208 uArgArgProAlaGlyGlnArgGlyAspGlyGlyProProGly.Leu 224
499 ATCAAGACGCGCGCGAGATTTCAAACGCGCGCTGTGATGACCG 548
225 ValHisGln.....ArgHisSerGlyAlaHisLeuGlyL 236
549 TTGACCGCAAGCAAAATTCATGTTGTGAAGCGACTGCGCGACGCTG 598
236 yGlnLeuAspArgValAlaGlyValAlaLeuValAlaGlnAlaProVal 253
599 CGTCGAAATGCTGCCACATGCAACATGATTCGGCGCGCGCAT 648
253 aGlnGly..... 255
649 CGTGGCGTTGAGTGGCAGC...CACATTCAT...TTCTGACGCGCG 692
256 ...LeuGlyLeuValAlaValProHisValHisValValAlaGlyProAl 271
693 CGCGCGCAATAAACCGTGTGACCATCAATATCAATGATTAATACCA 742
271 aGlyArgGlnAspArgAlaArgProAlaArgGlyHisGlyAspProAlaVal 288
743 TTGGCGCTTTGTTGCAACAGCGCGTGTGAACACCGCGCGATTTGCC 792
288 aAlaLeu.....GlnHisArgAlaAspHisProValAla 299
793 CTAGGTGTTCTCAAGTCAACAACCGCGCTGTG..... 828
300 Leu...GlyAspGlnLeuAspGlnArgGlyLeuValLarProAspArgAla 315
829 ...CGTACCGTTTGGGTGCGAAGATTCGAAATTAATCTGCGG 868
315 aGlyAlaGlnArgAlaValAlaGlnArgProGlyArgGlyLarGlyProAlaG 332
869 GCGAATGTTGACACACAGCAACCGCGATTCGCGTGGTGGTATTCAC 918
332 lylGlnValAlaAlaAspHisHis.....AlaAlaAspAlaLeuSer 346
919 GCGCGCATTTACACAAGCGCGACGATTAATTTGGACGCGCTACCAATCA 968
347 GlyAlaAlaGlnAspArgAlaGlnAlaLeuAlaGlyLeuAlaArgAspG 363
969 GATTC.....CGTTATGCA..... 983
363 nValHisProLeuValValArgThrGlyAspArgHisArgAspArgCysL 380
984 ...AGAAGCGCGAGCAA.....AGA 1001
380 euAspAspAlaArgProGlnGlnArgProGlyLeuAlaGlnHisArgArg 396

```

```

1002 GCTGTGCGCTGGTGGCGCGCGCAACAAATATCATCATCAGCGGTA 1051
397 ValGlnArgLeuAlaAlaAspAlaAla..... 405
1052 CAACCGTCGCGCATTTCTCGAAGAAACAAATCTTCAAGT..... 1091
406 ...ProArgGlyValAlaAlaArgGlnLeuArgValValAlaGlyAla 421
1092 ...CAMACAGCGCTACAGCGCGGCA...CGCGCGCAGCGTGC 1130
421 lAlaArgProHisGlnLeuGlnArgArgArgProGlnHis..... 435
1131 GATGTGACTTACAGCGCGGTGCTGCA..... 1163
436 ...ProAspGlyLeuGlyAlaValProGlnGluArg 446
1164 .TATCCTGCC.....CACCTGCTTTTGGCGGATTA 1194
446 gluProAlaGlyProSerGlyArgArgHisProProProProArgGlyA 463
1195 ATGCTGCGGATACGA.....CAGCGCGACGATGGGTTG 1232
463 rSerArgArgAlaArgAlaCysArgAlaSerArgAlaSerValSerAla 479
1233 ...CTGCAATTGACAGCAAGAACCT 1256
480 GlyLeuAlaAlaSerArgArgArgPro 488

```

seq_name: sp_human:Q96HA1

```

seq_documentation_block:
ID Q96HA1 PRELIMINARY; PRT; 984 AA.
AC Q96HA1;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE UNKNOWN (PROTEIN FOR MGC:3792).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LYMPH, AND LYMPHOMA;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC008794; AAH08794.1;
SQ SEQUENCE 984 AA; 99036 MW; 36DEDB4419EBA86 CRC64;

```

alignment_scores:

Quality:	Length:
Ratio: 0.525	Gaps: 21
Percent Similarity: 45.928	Percent Identity: 21.946

alignment_block:

US-09-303-518D-125/rev x Q96HA1 ..

Align seg 1/1 to: Q96HA1 from: 1 to: 984

```

1313 TTGGCCACACAGCGCGCTATTCGATTTGCGCGGCGAGACGAACTGCA 1264
||| :::::::::::::: ::::::::::::::
374 LeuSerGlnSerGlyProProGlyLeuLeuPro..... 384
1263 CAACGCGAGCTTTGTTGTCATTCACACCAACCAATGCC...TGCG 1217
||| ||| :::::::::::::: ::::::::::::::
385 .....SerProSerPheAspSerLyProProThrThrLeuLeuG 398
1216 CGCTGCGGTATCGCGCGACGATTAATGCGCAAAAGCGAGGTGGCAGG 1167
||| ::::::::::::::
398 lylLeuAlaProAlaProSerMet..... 405

```

1166 ATATCAAGGCGATACGCGCTCTGTAAGTACCAATCGGACACCATCGCGC 1117
406ValPro...AlaThrAspThrLy 412
1116 GTCCCGCCGCTTCAGCGCTGTGTGAACCTTGAAGAGCTTGTTTTTCAGA 1067
412ProThrLeuGlnAlaLutThrAlaThrLysProGlnAlaThrS 429
1066 AATGCCGAGGCTGTGACCGCGTGAAGAGATATTTGCGGTGGCGGCGCA 1017
429ProAlaProSerProAlaProLysGlnSerPheLeuPheGlyThrGlnAsn 445
1016 ACCGACCGCAACGCTCTTGGTGGGCGCTTCTTGATACGGAATCTG 967
446 ThrSerProSer.....ProAlaAla..... 453
966 ATTGTGTACGCTCCCAATAATCTGTGGCGCTTGTGTAATCGCGCT 917
454ProAlaAlaSerSerAlaProPheMetPheLysProI 466
916 TCATATCCGACGACGGGAATACGCGGTGTGTGTACCAATTCGCC 867
466 LeuPheThrAlaPro..... 470
866 GCAGTATTTGCGATCTTCGACCCCAAAACGGGTACGCAAGGCGCG 817
471ProLysSer.....GluLysGlu 477
816 TTGTGTGACTGAGAACCACTAGGCGCATACGCGCTCGGTTCAGAC 767
477 yProthr.....ProProGlyProSerValThrAlaThraAla..... 489
766 GCGCTGTGCACAAACAAAGCCCAATGTGTAATCATCTTGATATTGANG 717
490ProSerSerSerSerLeuProhrThr..... 498
716 GTTCACACGTTTATTTCGCCCGCACCGCTCGATCAATGAATGTCGCT 667
499ThrSerThrThrAlaProThr..... 505
666 GCCACTCAAAACG.....CGAGATCGGGCGCGCGCAATCATGTGTT 623
506PheGlnProValPheSerSerMetGlyProProAlaSer...Valp 520
622 CGATTTGGACGACATTTTCAGACGGCAGCTGTGCGCAGCTGCTTACAA 573
520ProAlaProPhePheLysGlnThrThrThrProAlaThrAlaPro 536
572 ACATGAGTTTGGCTGTGTCACAAAGCGTCAATACCAACAGGCCGCTT 523
537 Thr..... 537
522 GAAATCCTCGGCGCTCTTTGATTAATGACCGTAGGTCGCGACGACG 473
538ThrThrAlaProLeuPheThrGlyLeuAlaSerAlaThrSerA 552
472 GATTGTGTCTACGCAATGACGAATAATGCCACAGCGCTCGCATCG... 426
552 LaValAlaProIleThrSerAlaSerProSerThrAspSerAlaSerLys 568
425 ...ACGCGACGAATTTTGTGCAACGACGACGCGCCACGACGAGTCCACAA 379
569 ProAlaPheGlyPheGlyIleAsnSer...ValSerSerSerValSe 584
378 ACCGATGTGATCAGTTGCGGCGCACATCTTCGCCGCTTAAGTTGCCA 329
584 rTrhTrhTrhSerThrAlaThrAlaAlaSerIleProPheLeuPheGlyA 601
328 GC.....GCTTCAGTGGCTGACGCTTAAACTGCATTTGCTGCTG... 288
601 laProGlnAlaSerAlaAlaSerPheThrThrProAlaMetGlySerIlePhe 617
287CTTCAACGGCAATCAACGACTGACTGAAGTACGCGCTT 250

```

182 gProserIleTyrValIleAspAlaGlyAspIleIleGluTyrSHisA 199
160 ..... GGGCAA 165
199 rGProLysAspArgGlnLeuGlyLysLeuThrIleGlyThrSerGlyGlu 215
166 GTGCTGTTT ..... 174
216 IleAlaPheIleGlnSerThrMetSerAspArgGlyValValSerGlyAs 232
175 ..... GAAGACAAAACAT ..... 189
232 pValIleValLeuSerAspSerAspLysLysAsnLeuThrGlnGlyAsnG 249
190 ..... CCGGCGGTGTGTTTACTGCGCGGCTTCAGGCAAAATC 228
249 LuSerSerTyrSerHisIleValPhe ..... TyrAspGlyGluIle 262
229 GCGGCGCAT ..... CACGCTGGCGAAAGCGGCTACTTCAGTCAGTCGT 272
263 TyrValLeuGlnAsnHisGlnThrHisPheArgValIleSerLeuLeuTh 279
273 GATTGCGGTT ..... GAAGGACAGCAGAAATCGAGTTTGAACGCTAGCAGC 319
279 rArgLysValLeuThrPserGlyAspGlyValIleValTyrProAlaTyrSerP 296
320 CTGAA ..... GCGCTGGCAAACTTAAAGCGC 345
296 roSerPheHisIleSerSerGlyLysIleuAlaLeuAla...PheSerSer 311
346 GAAGAAAGTGGCGGCAACCTGATCCAAATCCGCTTGTGACGCTGCGTGC 395
312 ProAspGlnProGlnGlnValIleValLysAsnIleAspThrGlyValAl 328
396 CACCGTCGCTTCAGCAAAATCTCGCGCTGATGCGGCGGCTTCGCCA 445
328 uGlnArgSer ..... AlaI 333
446 TCTTCGTCATATGCGATGACACCAATCCGCTGCGCGACCTAGCGGC 495
333 LeuAsnSerGlyLeuLeuAsp ..... LeuLysAlaTyrProSerGln 346
496 ATTATCAAGAAAGCGCGGAGATTTCAAACGCGGCGCTGTGGTATTGAG 545
347 IleValIleGlnTyrPylAspAlaSerAspGlyLysIleTyrGlyIleLeuAr 363
546 CCGTTTGACGACGCAAAATCCATGTTTGAAGCAGCTGGCGGACAGC 595
363 gThrLeuAspProLys ..... 368
596 TGCCCTCGAAATGCTGCTCCAAACATCGAAACATGATTCGGGCGCG 645
369 ..... AsnProLeuIleValIleTyrValHis ..... GlyGlyPro 379
646 ..... CATCTCGCGGTTTGAAGTGCAGCAGCATTCATTTCAT 683
380 ThrSerPheSerTyrGlyAlaPheLeuAspArgThrSerValTyr ..... 394
684 CCAAGCCGCGCGGCGGATTAACCGCTGACCATCATTAATGATG 733
395 ..... LeuGlyTyrGlyPheSerValPheMetProAsnHisTyrGlyLys 409
734 TAAATACATGCGCGTTTGTTCACACAGCGCGCTGACACCGGAGCGC 783
409 eTyrValIleGlnLeuGlyArgGlnTyrAlaGlnSerAsnIleGlyAsp ..... 423
784 GTGATTGCGCTTGAAGTGTTCCTCAAGTCAACAACCGCGCTCTTCGCGAC 833
424 ..... LeuGlyGlyMetAspPheGlnLys ..... Va 432
834 CGTTTGGTGGGAAAGTATCGCAATATCTGCGGCGAATGTGTCACA 883
432 ILeSerGlyIleLysTyrLeuGln...GlnSerGlyLys...IleAspT 447

```

```

884 CAGACACCGCGCTGATTCCGCT ..... 906
447 hTyrAsnIlePheIleThrGlySerTyrGlyGlyTyrMetSerAla 463
907 ..... TCGTATTGACGCGCGATTCACAA...GG 935
464 LeuAlaValMetLysThrAspIlePheAsnAlaSerValSerLeuPheG 480
936 CGCGCAGATTAATTTGGACGCTACACACATCAATTCGTTATCGAAG 985
480 yLeSerAspThrPheSer...PheHisGlyThrSerAsnLeuTyrGlu 496
986 AAGGCGCG...AGCAAGAGCTGTTCGCGTTCGCGCGCAGCGCGAC 1032
496 rPAspArgIleHisLeuAspAlaAspProTyrPserPheGlnLysTyrAsp 512
1033 AATATCTCATCAGCGCTACAAACCTGCGCATTTCTGAAAACAAACT 1082
513 ArgTyrSerProIleArg ..... IleLysArgLysP 523
1083 CTTCAGTTCAACACACAGCGGCTGACGCGCGCGCGCATGTGCGCA 1132
523 oLysThrProValLeuLeuMetHisGlyValAsnAspLysTyrValProI 540
1133 TTGCTACTTAC 1143
540 LeGlyGlnTyr 543

```

seq_name: sp_invertebrate:Q9BKV7

seq_documentation_block:

```

ID Q9BKV7 PRELIMINARY; PRT; 1325 AA.
AC Q9BKV7;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE PPG3.
GN PPG3.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Myler P.J., Sisk E., Cawthra J., Handley F., Vogt C., Robertson L.,
RA McDonald P., Ivens A., Nguyen D., Munden H., Stuart K., Worthey E.A.,
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC084329; AKK1375.1; -.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003591; LRR_tyr.
DR Pfam; PF00560; LRR; 6.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00369; LRR_tyr; 5.
SQ SEQUENCE 1325 AA; 133855 MW; 8898928BAEF9418F CRC64;

```

alignment_scores:

```

Quality: 106.00 Length: 438
Ratio: 0.500 Gaps: 13
Percent Similarity: 48.402 Percent Identity: 20.776

```

alignment_block:

US-09-303-518d-125/rev x Q9BKV7 ..

Align seg 1/1 to: Q9BKV7 from: 1 to: 1325

```

1337 TCCCTTCGAATGTTTCCAGCATTTGCGCAACGCGGCGGCTATTGCTA 1288
111 111 111 111 111 111 111 111 111 111 111 111 111
741 SerSerSerSerAlaProSerAlaSerSerSerSerAlaProSerSer 757
1287 TTTGCCCGGCGGACAGCAAGTGTGACAAAGAGTCTTCTTCGTCAT 1238
111 111 111 111 111 111 111 111 111 111 111 111

```



```

757 rSerSerAlaProSerAlaSerSerSerSerAlaProSerSerSers 774
1237 CCAAGCACCCCAATGCTGCGGCTGTCGATGCGCCAGCAATTAATCG 1188
1187 CGCAAAAGAGGGGTGGCAGATTCAGAGGCATCAGCGCTGCTGAAT 1188
791 AlAProSerAlaSerSerSerSerAlaProSerSerSerSerAla 807
1137 ACCAATCGCACCATGGCGGCTGCGCCCTTCAGCGCTGTGTGACT 1088
807 aProSerAlaSerSerSerSerAlaProSerSerSerSer 820
1087 TGAAGATTTGTTTTCAGAAATGCGCAGGTTGTACGGTGATGAG 1038
821 .....Ser 821
1037 TATTGTCGGCTGCGGCGCACCCAGCAAGCTCTTTGCTGCGGCC 988
822 AlAProSerAlaSerSerSerSerAlaProSerSerSerSerAlaPr 838
987 TCTTCGATACGGAATCTGATGTGTAGCGTCCCAATATCTGCTGC 938
838 oSerAlaSerSer.....SerSerA 845
937 CCGCTTGCTGTAATGCGCGCTTCATACCGACCGGAATCAGCGGTTG 888
845 lAProSerSerSerSerSerAlaProSerAlaSerSerSerSerAlaPro 861
887 TCTGTGTCACCAATTCGCGCGCAGTAATTTGGATCTTTGCGACCA 838
862 SerSerSerSerAlaProSerAla...SerSerSerAlaProSe 877
837 AACGTCAGCAAGAGCGCGTTGTTGACTTGAGAACCCAGCGGCA 788
877 rSerSerSerSerAlaProSerAlaSerSerSerSerAlaProSers 894
787 TCAGCGCGCTGCGGTCAGCGGCTGTCGAACCAACGCGCAATGTA 738
894 erSerSerSerAla.....ProSerAlaSerSerSerSerAlaPro 907
737 ATTACATCTGATTAATGATGTCACAGGTTTATTGCGCGCAGCGG 688
908 SerSerSer.....SerSerSerAlaProSerAl 917
687 CTGATGAATGATGCTGCGCTGCAACCGGAGAGATGCGGCGCG 638
917 aSerSerSer.....SerAlaProSerSerSerSerSerAlaPro. 931
637 CGAATTCATGTTGATGTTGGCAGCATTTTCAGAGCGCAGCTGCG 588
932 .....SerAlaSerSerSerSerAlaProSerSerSerSerAla 946
587 CCAAGCTGCTTCGAACATGATTTTGGCTTCGATCAACGCTCAATAC 538
947 ProSerAlaSerSerSer.....SerAlaProSerSerSerSe 959
537 CAACAGCGCGCTTGAATCTGCGGCGCTTCTTGTATGACCGTAG 488
959 rSerAlaProSerAlaSerSerSerSerAlaPro..... 970
487 GGTGCGCAGCGAGGATTTGTCATCGATTCAGCAAGATGCGGAC 438
971 ..SerSerSerSerSerAlaProSerAlaSerSerSerSerAlaProSer 986
437 GGTGCGCATGACGCGCAGGAATTTGTCGACGAGCGGTCGCGACGC 388
987 SerSerSerSerAla..... 992
387 AGTCACAACCGGATTTGATCGAGTTCGCGCACTTCTGCGCGCTTA 338
993 .....ProSerAlaSerSerSerSerAlaProSerSerSers 1006
337 AGTTTGCCAGCGCTTCAGGTGGTAGCGTTCAAACTGATTTGCTG 288
1006 eAlAProSerSerAlaSerSerSerSerAlaProSerSerSerSerAla 1022
287 CCTTCAGCGGCATACGACTGACTGAGTACGCGCTTTGCGCCAGGTG 238
1023 ProSerAlaSerSerSerSerAlaProSerSerSerSerSer..... 1036
237 ATTCGCGCGATTTTGCTGAAGCGCGCGCAGTAACACACGCGCGAT 188
1037 .....AlAProSerAlaSerSer..... 1046
187 TCTTTTGTCTTCAACAGCACTTGCCCTTTTTCAGCGCATGCGCTTC 138
1047 .....SerSerSerSerSerAlaProSerAlaSerSerSerSerAla 1060
137 TTGACTTTCATCGAGGCGCATACCGGATATTTTCGCCAACAACGC 88
1061 ProSerSerSerSerSerSerAlaProSerSerAlaSerSerSerSerAlaPr 1077
87 GACTTCGTAATGCGCGGCGCTGTAACGCGCTTCTCGGCTGCTCCG 38
1077 oSerSerSerSerSerAlaProSerAlaSerSerSerSerAlaProSers 1094
37 CGATGGCAGGTTT 24
1094 erSerSerSerPhe 1098

```

